



The Yukon Electrical Company Limited
An **ATCO** Company

October 31, 2013

Yukon Public Utilities Board
Box 31728
Whitehorse, YK Y1A 6L3

Attention: Bruce McLennan
Board Chair

Dear Mr. McLennan:

Re: Procedural Direction – Use of Hyperlinks in documents filed in proceedings

Further to the Procedural Direction dated October 4, 2013, Yukon Electrical is providing PDF copies of all hyperlinked documents referred to in its 2013-2015 General Rate Application and related Responses to Information Requests, other than those provided previously.

We trust the foregoing is satisfactory. If you have any questions in respect of the above, please contact me at (780) 733-2489.

Yours truly,

Original Signed by

James Grattan, CA
Director, Regulatory



The Yukon Electrical Company Limited
An **ATCO** Company

Attachment	Reference	Description	Hyperlink
Attachment 1	Application, page 11-5	Examples of Spurious Regression	http://www.eco.uc3m.es/~jgonzalo/teaching/timeseriesMA/examplelesspuriousregression.pdf
Attachment 2	Application, page 11-11	Manitoba Hydro 2012 Electric Load Forecast	http://www.hydro.mb.ca/regulatory_affairs/electric/gra_2012_2013/attachment_1.pdf
Attachment 3	Application, page 11-13	Weather Normalization Presented by J. Stuart McMenamin	www.aeic.org/load_research/docs/1_WeatherNorm_McMenamin.ppt
Attachment 4	Application, page 11-15	Government of Alberta 2010-2013 Economic Outlook (page 118)	http://www.finance.alberta.ca/publications/budget/budget2010/economic-outlook.pdf
Attachment 5	Response to Information Request CW-YECL-1(a)	Canada Mortgage and Housing Corporation Northern Housing Report, release in 2012	http://www.cmhc-schl.gc.ca/odpub/esub/65446/65446_2012_A01.pdf?lang=en
Attachment 6	Response to Information Request UCG-YECL-19(d)	The International Group of Liquefied Natural Gas Importers, LNG Information Paper No.1	http://www.giignl.org/fileadmin/user_upload/pdf/LNG_Safety/1-LNG_Basics_8.28.09_Final_HQ.pdf
Attachment 7	Response to Information Request UCG-YECL-104(a)	British Columbia Ministry of Energy and Mines, Liquefied Natural Gas, A Strategy for B.C.'s New Industry	http://www.gov.bc.ca/ener/popt/down/liq uefied_natural_gas_strategy.pdf
Attachment 8	Response to Information Request UCG-YECL-128(a), Attachment 2, Page 9 of 20	Dollars to Sense courses	http://oee.nrcan.gc.ca/industrial/training-awareness/2442
Attachment 9	Response to Information Request UCG-YECL-128(a), Attachment 2, Page 9 of 20	Energy Benchmarking tools	http://oee.nrcan.gc.ca/industrial/technical-info/benchmarking/4377#energy-performance
Attachment 10	Response to Information Request UCG-YECL-128(a), Attachment 2, Page 9 of 20	Reading your bill	http://www.yukonenergy.ca/customer/tips_tools/reading/

Serving the Yukon since 1901



The Yukon Electrical Company Limited
An ATCO Company

Attachment	Reference	Description	Hyperlink
Attachment 11	Response to Information Request YUB-YECL-26(a), Attachment 1, Page 9 of 83	The California Evaluation Framework, Prepared for the California Public Utilities Commission and the Project Advisory Group, June 2004	http://www.calmac.org/publications/California_Evaluation_Framework_June_2004.pdf
Attachment 12	Response to Information Request YUB-YECL-26(a), Attachment 1, Page 9 of 83	California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals, Prepared for the California Public Utilities Commission, April 2006	http://www.calmac.org/events/EvaluatorsProtocols_Final_AdoptedviaRuling_06-19-2006.pdf
Attachment 13	Response to Information Request YUB-YECL-26(a), Attachment 1, Page 10 of 83	International Performance Measurement and Verification Protocol, Concepts and Options for Determining Energy and Water Savings, Volume 1, January 2012	http://www.evo-world.org
Attachment 14	Response to Information Request YUB-YECL-26(a), Attachment 1, Page 18 of 83	California Standard Practice Manual, Economic Analysis of Demand-Side Management Programs and Projects, October 2001	http://www.energy.ca.gov/greenbuilding/documents/background/07-J_CPUC_STANDARD_PRACTICE_MANUAL.PDF
Attachment 15	Response to Information Request YUB-YECL-42(a)	Sproule Forecasts	http://www.sproule.com/forecasts
Attachment 16	Response to Information Request YUB-YECL-42(f)	LNG for Yukon Energy Power Generation, A Life Cycle Emissions Inventory, July 2013	http://www.yukonenergy.ca/downloads/db/1265_LNG%20LCA%20Pembina%20Report.pdf
Attachment 17	Response to Information Request YUB-YECL-56, Attachment 1, Page 10 of 33	Production of Canada's Leading Minerals, Natural Resources Canada, December 2011	http://mmsd.mms.nrcan.gc.ca/stat-stat/prodprod/2011-eng.aspx

Serving the Yukon since 1901

Examples of Spurious Regression

(“Regression that does not make any sense.”)

Typical symptom: “High R^2 , t-values, F-value, but **low D/W**”

1. Egyptian infant mortality rate (Y), 1971-1990, annual data,
on Gross aggregate income of American farmers (I)
and Total Honduran money supply (M)

$$\hat{Y} = 179.9 - .2952 I - .0439 M, \quad R^2 = .918, \quad D/W = .4752, \quad F = 95.17$$

(16.63) (-2.32) (-4.26) Corr = .8858, -.9113, -.9445

2. US Export Index (Y), 1960-1990, annual data,
on Australian males' life expectancy (X)

$$\hat{Y} = -2943. + 45.7974 X, \quad R^2 = .916, \quad D/W = .3599, \quad F = 315.2$$

(-16.70) (17.76) Corr = .9570

3. US Defense Expenditure (Y), 1971-1990, annual data,
on Population of South African (X)

$$\hat{Y} = -368.99 + .0179 X, \quad R^2 = .940, \quad D/W = .4069, \quad F = 280.69$$

(-11.34) (16.75) Corr = .9694

4. Total Crime Rates in the US (Y), 1971-1991, annual data,
on Life expectancy of South Africa (X)

$$\hat{Y} = -24569 + 628.9 X, \quad R^2 = .811, \quad D/W = .5061, \quad F = 81.72$$

(-6.03) (9.04) Corr = .9008

5. Population of South Africa (Y), 1971-1990, annual data,
on Total R&D expenditure in the US (X)

$$\hat{Y} = 21698.7 + 111.58 X, \quad R^2 = .974, \quad D/W = .3037, \quad F = 696.96$$

(59.44) (26.40) Corr = .9873

2012 ELECTRIC LOAD FORECAST

(For External Use)

MARKET FORECAST
MAY 2012
APPROVED JULY 2012



ASSUMPTIONS

Economic Assumptions

Economic forecast assumptions are taken from the 2012 Economic Outlook and the 2012 Energy Price Outlook. These documents contain Manitoba Hydro's forecasts of economic variables including prices of electricity, natural gas and oil, Gross Domestic Product (GDP), Manitoba population and residential customers.

The following are the economic variables used in the preparation of this Electric Load Forecast:

Residential Customers - The number of Residential Basic customers in Manitoba is forecast to increase by 1.2% (5,581 units) in 2012/13 and averages 1.2% per year over the forecast period. This compares to a historical average increase of 0.9% per year over the last ten years. This is used in the Residential and GS Mass Market customer forecasts.

Electricity Prices - The electricity price forecast is based on CPI and rate increase projections contained in the Integrated Financial Forecast. The nominal electricity price is forecast to increase by 2.0% in 2012/13, increase 3.5% per year until 2023/24 and then increase by 2.0% per year throughout the rest of the forecast period. This is used in the Residential customer forecast.

Natural Gas Prices – Manitoba Hydro views the natural gas price forecast as commercially sensitive information. Consistent with the Clean Environment Commission and Electric General Rate Application, this information will not be publicly disclosed. This is used in the Residential customer forecast.

Gross Domestic Product (GDP) - Real economic growth in Manitoba is forecast to be 2.3% in 2012/13. It is expected to stay at approximately that level for several years, then decline to 1.7% by 2019/20 and remain at that level for the remainder of the forecast period. This is used in the GS Mass Market customer forecast.

Normal Weather Assumptions

Weather for forecast purposes is measured by degree days. Winnipeg temperatures are used, as Winnipeg is central to most of the weather-dependent load (Residential and General Service Mass Market) in Manitoba.

Cold weather is expressed in Degree Days Heating (DDH), which is the number of average degrees colder than 14 degrees Celsius each day. Hot weather is expressed in Degree Days Cooling (DDC), which is the number of average degrees warmer than 18 degrees Celsius each day. Daily temperature is the average of the high and low temperature for the day. The equations are:

$$DDH = \text{sum} (\max(0, 14 - (\text{Daily high} + \text{Daily low}) / 2))$$

$$DDC = \text{sum} (\max(0, (\text{Daily high} + \text{Daily low}) / 2 - 18))$$

The base temperature of 14 degrees for DDH is the temperature below which most buildings have their heating systems (furnaces) running.

The base temperature of 18 degrees for DDC is the temperature above which buildings start to run their space cooling systems (air-conditioning).

The forecast is prepared assuming normal weather. Normal weather is determined from the 25 year average of Degree Days Heating and Degree Days Cooling in Winnipeg over the period April 1987 to March 2012.

The 25 year weather normals used for every year of this forecast are 4,518.4 DDH and 188.1 DDC. This is a decrease of 18.3 DDH from last year's normal of 4,536.7 DDH, and an increase of 3.9 DDC from last year's normal of 184.2 DDC.

The range of DDH from 1987 to 2012 was from a warm winter of 3,677.6 DDH in 2011/12 (840.8 DDH below normal) to a cold winter of 5,439.3 DDH in 1995/96 (920.9 DDH above normal).

The range of DDC from 1987 to 2012 was from a cool summer of 71.8 DDC in 2004/05 (116.3 DDC below normal) to a hot summer of 364.1 DDC in 1988/89 (176.0 DDC above normal).

Demand Side Management (DSM) in the Forecast

This forecast reflects future DSM savings associated with improved equipment efficiency standards and existing Provincial building codes. This is the only DSM initiative that is specifically accounted for in the forecast.

Savings due to DSM programs to date are embedded in the historical data that is the basis for this forecast. The current level of past achieved DSM savings is assumed to remain in place throughout the future. Future DSM savings arising from future Power Smart offerings and market engagement above the current level and incremental to the above mentioned standards and codes are treated as a supply-side resource and are not included in this document. They are accounted for separately in Manitoba Hydro's Power Smart Plan and Power Resource Plan.

For customers involved in DSM initiatives such as the Load Displacement and Alternative Energy, the most recent years of history are used as the basis of their forecast. Any increase or loss to their alternative energy supply will not be included in the forecast, and are accounted for with the other DSM initiatives within the Power Smart Plan.

METHODOLOGY

Residential Basic Methodology

The Residential Basic forecast was calculated using a detailed end use approach. The forecast of the total number of Residential Customers was from Manitoba Hydro's 2012 Economic Outlook. The 2009 Residential Energy Use Survey provided current end use saturation rates, detailed information on newly constructed dwellings, and appliance age distributions and their expected lifetimes. The end use assumptions include usage information and efficiency improvement information. The number of appliances and their estimated usage were multiplied together to calculate an energy forecast for each end use, and then all uses were combined to calculate the total use for the Residential End Use Forecast.

a) Total Number of Residential Customers - The Economic Analysis Department forecast the total number of residential customers for the 2012/13 to 2031/32 period. This customer forecast was the primary input for the Residential End Use Model.

b) Customers by Dwelling Type and Area - The 2009 Residential Survey was used to estimate the number of customers to various dwelling types (Single Detached, Multi-family Attached, and Individually-Metered Apartments). Single detached dwelling types are sub classified as Winnipeg, Gas Available Areas Outside of Winnipeg, and Gas Unavailable Areas.

c) Customers by Heating Type - Each combination of Dwelling Type and Area are divided into two groups: Electric Heat Billed customers and Other Heat customers. Electric Heat Billed customers pay for their space heat with their electricity bill. Other Heat customers may use natural gas or propane, or may use electric heat but are not billed directly.

For Single Detached Homes in Gas Available Areas Excluding Winnipeg, the number of newly constructed homes choosing electric heat was econometrically forecast using the following equation:

Change in Percentage of Newly Constructed Single Detached Homes in Gas Available Areas Outside of Winnipeg with Electric Heat Billed (t)

$$= 0.001 + 0.668 \times \text{Chg PG/PE}$$

Change in PG/PE

$$\begin{aligned} &= \text{Price of Gas per mMBTU (t-1)} / \text{Price of Electricity per mMBTU (t-1)} \\ &\quad - \text{Price of Gas per mMBTU (t-2)} / \text{Price of Electricity per mMBTU (t-2)} \end{aligned}$$

R-squared: 44.6%

T-stats:

Constant	: 0.14
Chg PG/PE	: 3.81

The same model with different coefficients was used to forecast heating appliances in newly constructed single detached homes in Winnipeg.

d) Appliance Forecast - Historical saturation and age distribution data was collected from the 2009 Manitoba Hydro Residential Survey. Saturations were forecast using a birth/death/replacement model.

e) Average Appliance Usage - The current estimates of annual appliance usage, also called Unit Energy Consumptions (UECs) were calculated using Residential Survey information, Conditional Demand Analysis techniques, and expert opinion. The survey results were screened for consumption records and survey completeness. Missing values for the size of home, people per household and income questions were imputed. Degree days heating/cooling and demographic factors such as income and people per household were added to help explain usage variations. They were then normalized for the average customer.

f) New Appliance Usage - New end uses are typically more efficient than existing stock. UECs for new appliances were calculated and these were used in the birth/death/replacement model.

g) Total Energy Use - The forecast number of appliances was multiplied by the forecast UECs to get the forecast kW.h per appliance. The appliance usages were summed to get the total use for the Residential Basic rate class. Efficiency improvements due to future standards that were not handled by the birth/death/replacement model were handled separately. Two-thirds of the Plug-In Electric Vehicle forecast was added to the Residential GW.h. (The other one-third of the electric vehicle forecast was allocated to the General Service Mass Market forecast.)

General Service Mass Market Methodology

There are four rate classes in the General Service Mass Market sector:

- i) Small Non-Demand (0 to 50 kV.A),
- ii) Small Demand (50 to 200 kV.A),
- iii) Medium (above 200 kV.A but do not own transformation capabilities), and
- iv) Large (above 200 kV.A and own their own transformation capabilities).

Total GW.h for the General Service Mass Market sector is forecast by multiplying the forecast number of customers in each rate class by the forecast average use per customer in each rate class. The total number of customers in the GS Mass Market sector is forecast using econometric models, and then the total is allocated among the rate classes based on historical data. The average use per customer is forecast using the five year average for each rate class.

One-third of the Plug-In Electric Vehicle forecast is added to the GW.h for the Small Non-Demand rate class. (The other two-thirds of the electric vehicle forecast are allocated to the Residential forecast.) Forecast savings from future standards and construction codes are taken off of the forecast to calculate the Total Use.

General Service Mass Market Customer Forecast

Econometric analysis of sales data is used to develop models for the number of customers. Forecasts of Manitoba GDP and Manitoba Hydro Residential Customers from the Manitoba Hydro 2012 Economic Outlook are then input into the models, which generate forecasts for the number of customers for each year of the forecast period.

The number of customers at fiscal yearend was forecast using the following calculations for each year (t):

Number of Customers (t)

= Number of Customers (t-1)

+ Change in the Number of Customers (t)

Change in the Number of Customers (t)

= Number of Customers (t-1)

x Percentage Change in Number of Customers (t)

The percentage change in number of customers was modeled using yearend historical customer data from 1984/85 to 2011/12. The resulting model and parameters are as follows:

Percentage Change in Number of Customers (t)

= -0.002 + 0.135 x CGDP + 0.629 x CRES

CGDP - Annual Percentage Change in Manitoba Gross Domestic Product

CRES - Annual Percentage Change in Residential Basic Customers

R-squared: 54.5%

T-stats:

Constant : -1.27

CGDP : 3.55

CRES : 3.44

General Service Mass Market customer growth is allocated to Medium and Large classes using their 10 year average percentage of Mass Market customer growth, the Small Non-Demand class is allocated 10% of new customers, and the rest of the growth is allocated to the Small Demand class.

General Service Mass Market Average Use Forecast

The average use for each rate class is forecast to remain at its five year average over the forecast period. Specifically these are:

- i) Small Non-Demand - 31,027 kW.h per year
- ii) Small Demand – 166,454 kW.h per year
- iii) Medium - 1,582,552 kW.h per year
- iv) Large - 5,877,414 kW.h per year

Top Consumer customers are excluded from these classes for forecasting purposes.

Customers are assigned to a rate class depending on their usage. If usage by an individual customer increases (or decreases) sufficiently then they will be re-assigned to the appropriate rate class. These shifts tend to offset each other over time so individual classes have not shown significant upward or downward trends in average use. By definition, the truncation of these classes results in relatively stable average use for each class.

General Service Mass Market Total Use Forecast

Total Use (t)

- = Number of Small Non-Demand Customers (t)
 - x Average Annual Use of Small Non-Demand Customers (t)
 - + 1/3 of Plug-in Electric Vehicle Forecast
- + Number of Small Demand Customers (t) x Average Annual Use of Small Customers (t)
- + Number of Medium Customers (t) x Average Annual Use of Medium Customers (t)
- + Number of Large Customers (t) x Average Annual Use of Large Customers (t)
- Forecast Savings from Construction Codes (t)

General Service Top Consumers Methodology

Top Consumers is made up of the largest electricity users of Manitoba Hydro. The general criterion is that a company needs to have used 50 GW.h in a year, or have the potential to consume 50 GW.h in a year. A Top Consumer is not necessarily located in one place, but may consist of services at number of locations throughout the Province. A Top Consumer will be one company, but may count as multiple billing customers.

Each company in the Top Consumers group is forecast individually. Information on individual company operating plans is collected from industry news, Manitoba Hydro's economic experts and Manitoba Hydro's Key & Major Account representatives. This information is used to prepare company specific forecasts.

Normally, information is only available over the next 3 to 5 years for any company. These short term considerations are taken into account, and then the company's forecast remains constant.

To account for longer term growth in this group of consumers, a special classification called Potential Large Industrial Loads (PLIL) has been created. PLIL is used instead of attempting to forecast each consumer individually for the long term. It represents the natural growth of all the top consumers as a group, as well as unexpected major expansions, new customers, or loss of customers from GS Top Consumers.

Starting in 2014/15, 100 GW.h a year is forecast for PLIL to account for unforeseen expansion, contraction and growth. This will result in the addition of 1,700 GW.h of PLIL after 20 years.

Electric Vehicles

The methodology for forecasting Electric Vehicles was to research relevant recent literature and to apply appropriate assumptions from this literature to Manitoba's situation. Historical data on automobile registrations per year in Manitoba was used to help estimate future trends. The forecast section on Electric Vehicles provides the details.

Other Sectors

Seasonal, Water Heating, Lighting

Most of the smaller sales sectors, including Seasonal, Flat Rate Water Heating and Area and Roadway Lighting were done by analysis of changes in the number of customers or services and in changes in average use per customer or service. Growth rates were applied based on history and a best estimate as to what the future will bring.

Diesel

Each of the diesel towns was individually forecast. An additional forecast was produced assuming that the customers would be converted to the Integrated System and given 200 amp service which would allow electric heating.

Monthly Sales Allocations

Monthly percentages of customer growth through the year and GW.h for the month of the year were averaged for the past five years. These were then applied to the forecast annual customers and kW.h to get the monthly forecast.

Monthly System Energy and Peak

An hourly load model was developed using econometric modeling techniques to determine the hourly load shape of the system. This load shape was used to forecast monthly system energy, annual system peaks, and monthly system peaks.

Data used in the hourly load model included hourly data for Common Bus, Top Consumers, and the larger GS Mass Market customers. Load research sample data was used to estimate Residential and smaller Mass Market customers. Twenty-five years of hourly temperature data were used to simulate the hourly load shape for each sector.

The modeled load shape of each sector was applied to the annual energy forecast to determine each sector's contribution to peak demand. The sectors were summed to get an hourly forecast of General Consumers Sales. Distribution Losses were calculated as a percentage of sales using the five year average of annual distribution losses. Transmission losses were calculated for each month as a percentage of Common Bus using the five year average for each month.

General Consumers Sales, losses, and miscellaneous loads were added to arrive at the forecast hourly system load for Manitoba. The maximum hourly loads in each month were used to estimate the monthly System peak and annual System peak.

Electric / Gas / Water

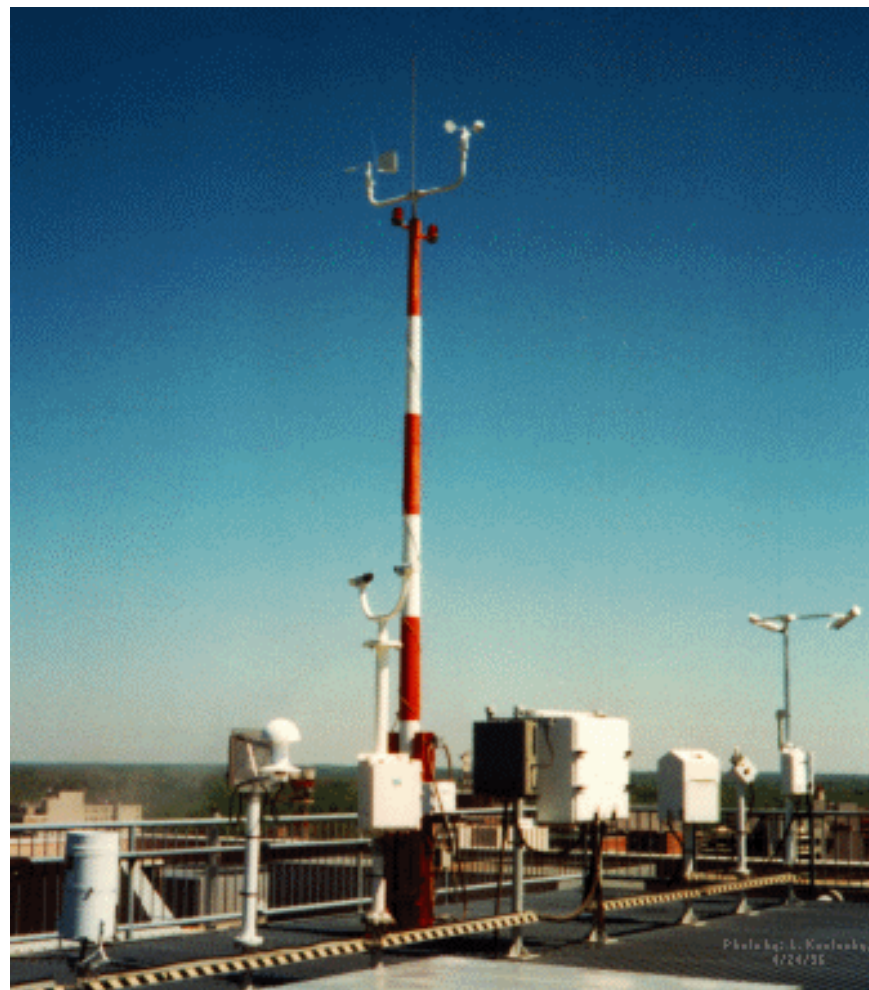


Weather Normalization

**Presented by
J. Stuart McMenamin**

Agenda

- Overview of Weather Normalization Concepts
- Itron Benchmarking Study
 - Uses of Normalization
 - Definition of normal weather
 - Cold weather variables
 - Hot weather variables
 - Weather adjustment methods
- Review of study results
 - Discussion of issues
 - Presentation of industry practices
 - Recommendations
- Conclusion

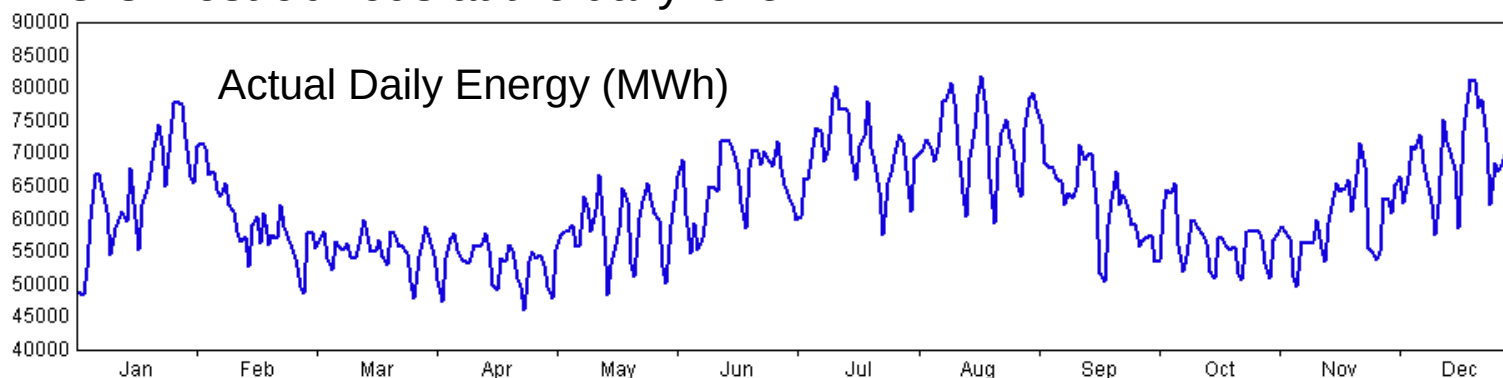


Applications of Normal Weather

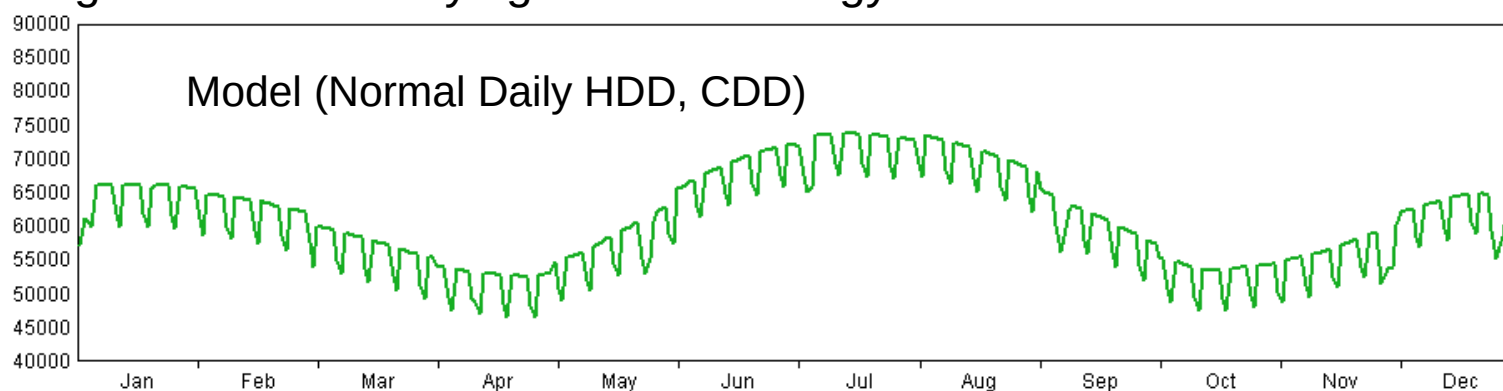
- Financial forecasting with normal weather
 - Budget forecast – usually calendar month energy use Monthly, Averaged
 - Budget forecast of daily energy (for daily tracking) Daily, Averaged
 - Monthly peak forecast Daily, peak
 - Annual peak forecast Daily, peak
- Financial variance analysis – adjusting from actual to normal
 - Adjust billed monthly energy use by class Monthly, Averaged
 - Adjust booked (calendar month) energy use by class Monthly, Averaged
 - Adjust system energy and peak demands Daily, peak
 - Adjust daily system energy to track against budget Daily, Averaged
- Facility Planning – weather scenario or simulations
 - Generate 8760 hour load forecast for dispatch simulations Hourly, Chaotic
 - Design day forecast Daily, Extreme
 - Simulate peak distributions, coincidence factor distributions,... Daily, Multiple
- Market Analysis
 - Generate a “typical” load profile for a customer Hourly, Chaotic
 - Generate facility baselines for budgets, pricing, ... Monthly or Daily

Why Weather Normalization is Useful

- Actual energy use is chaotic. The main source of irregularity is weather. This is most obvious at the daily level



- Models using normal weather create order out of chaos, providing insights about underlying trends in energy use and revenues.

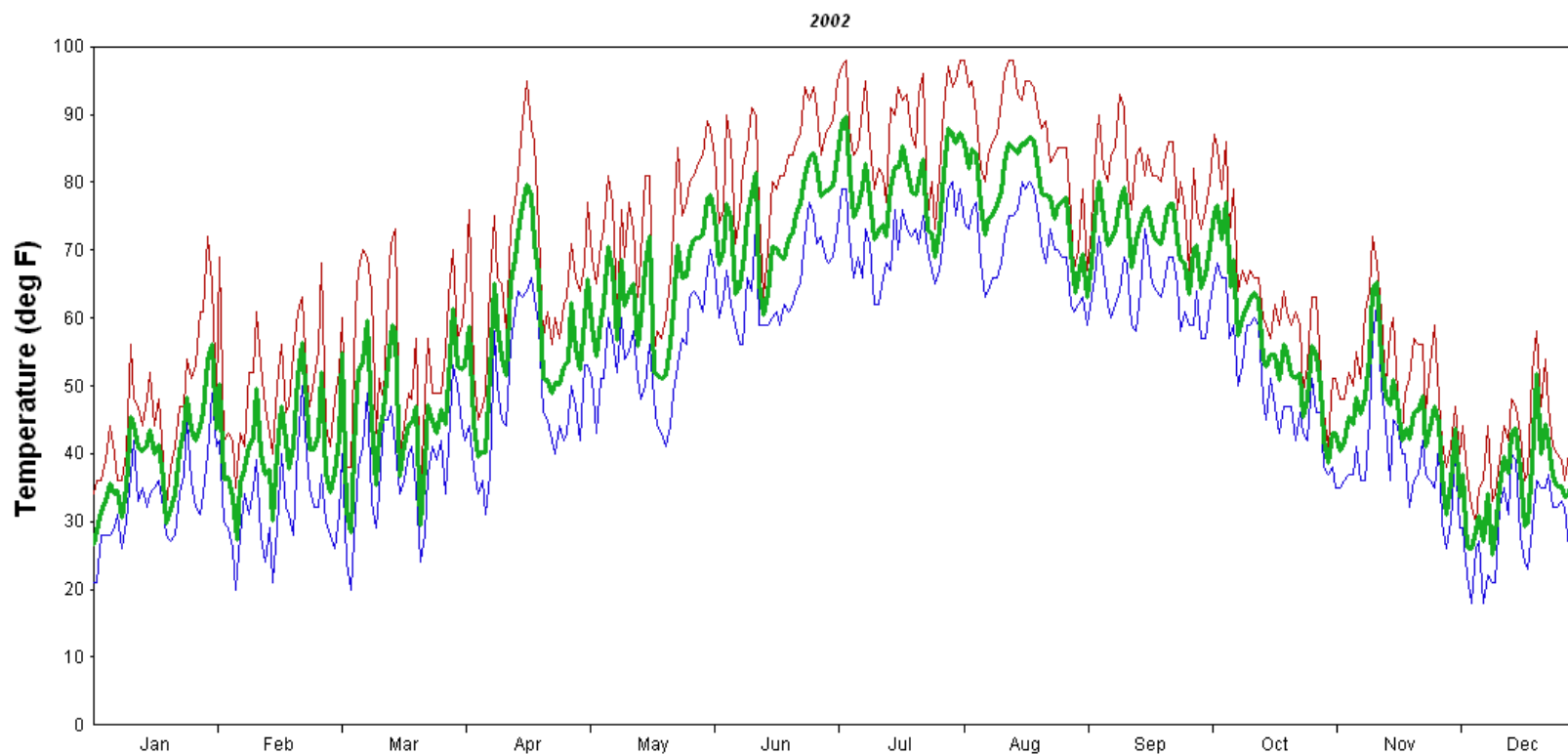


Types of Normal Weather

- Chaotic scenarios (daily or hourly)
 - Selected actual weather patterns (e.g., TMY files)
 - Rank and average, assign to calendar days. Intended to have normal extreme days as well as normal averages
- Daily Normals
 - Averaged by date
 - Apply nonlinear operations (HDD, CDD) before averaging
- Monthly Normals
 - Built up from daily normals (HDD and CDD values)
 - Average of monthly values – calendar month and/or billing month
- Peak Producing Weather
 - Weather based – hottest, coldest – normal extremes or design day (1 in 10)
 - Load based – weather on historical peak days
- Simulation Based
 - Multiple years of actual weather simulated through weather response function

Actual weather data is chaotic

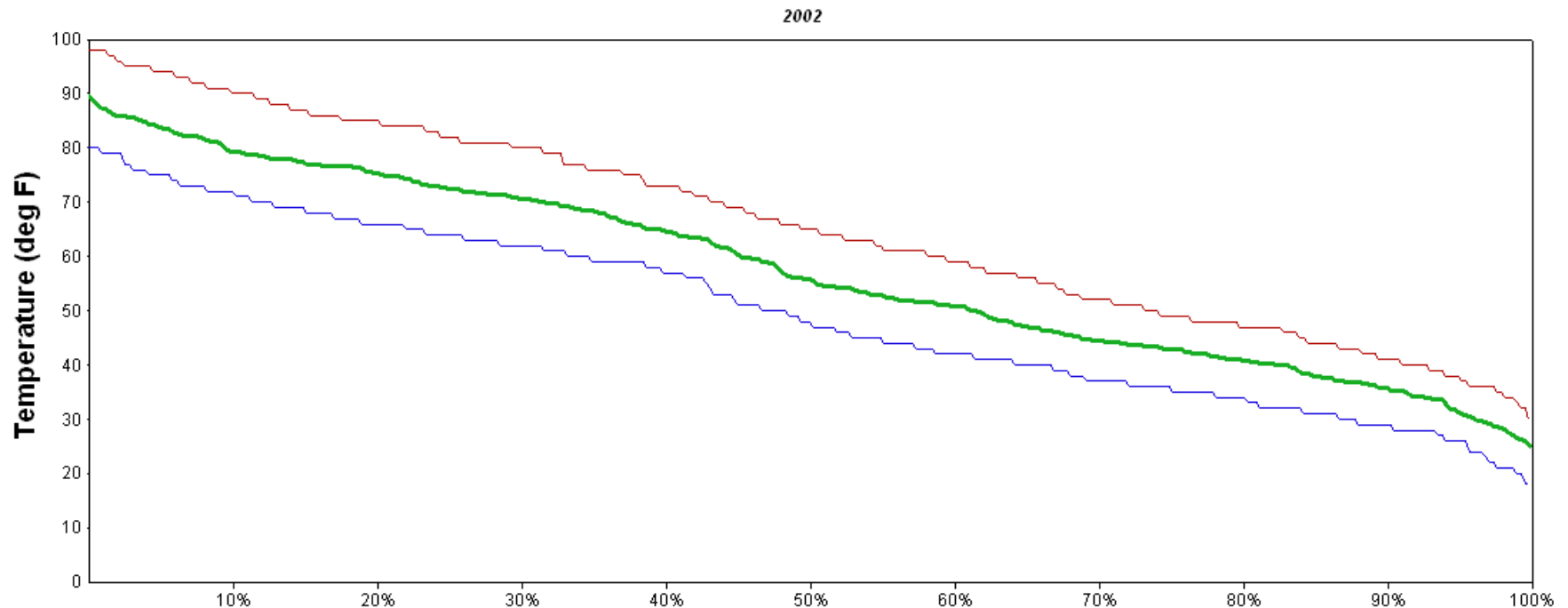
- Actual weather data is chaotic. Some years are less irregular than others. But all years are irregular. So...what is normal?



Graph shows daily high, low, and average values for one year.

Rank and average methods

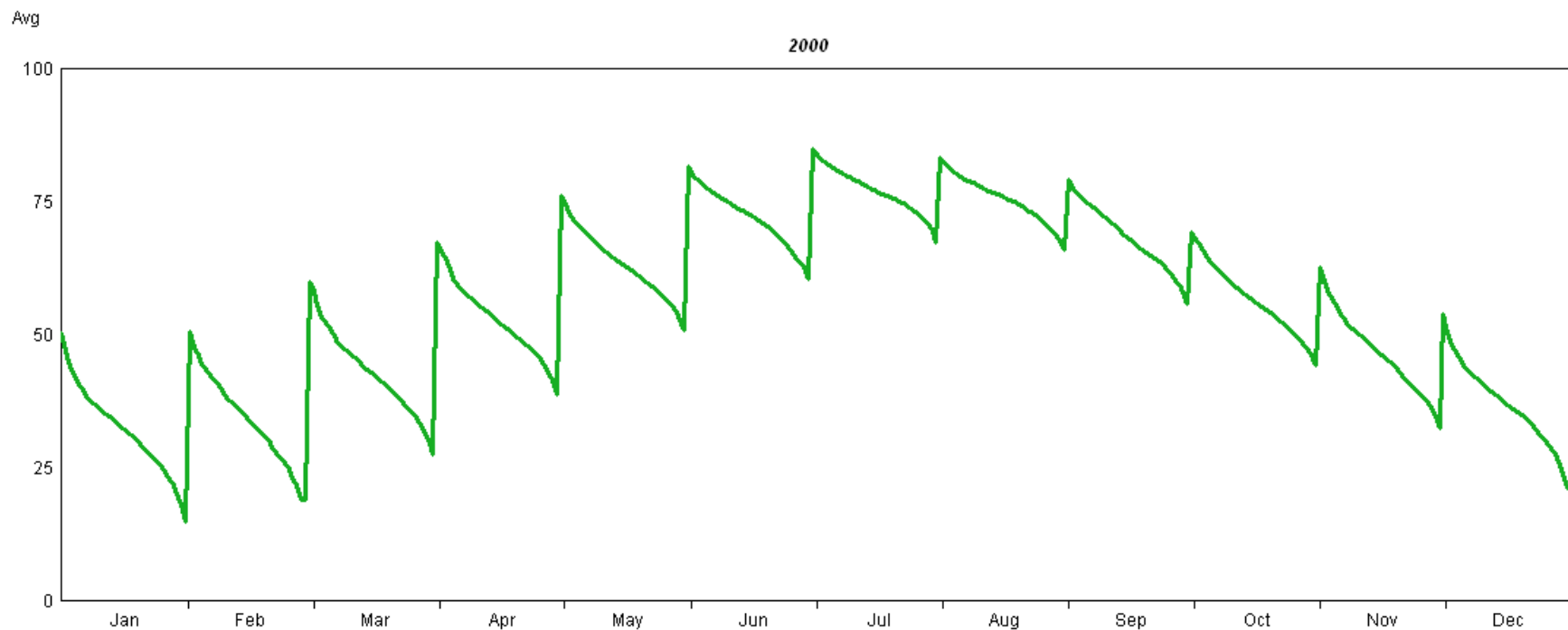
- Rank and average methods start by creating a “temperature duration” curve. The first number represents the hottest day in the year. Averaging across years gives a typical hottest day, second hottest day, ... , and on to a typical coldest day.



Graph shows sorted values for one year. High, low and average are sorted separately.

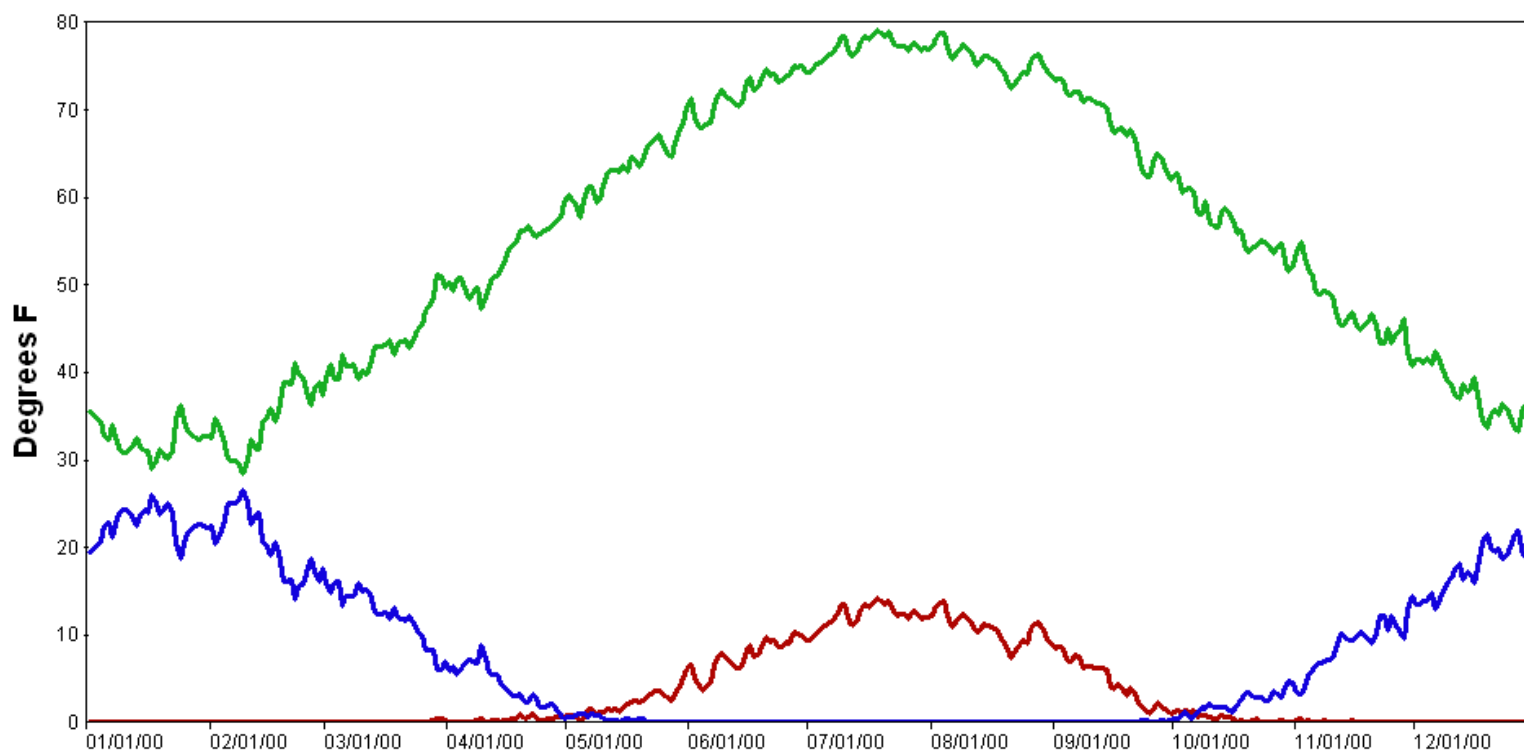
Rank and Average Methods

- Rank and average results have normal averages and normal extremes for each month (or season). They still need to be assigned out to specific dates, creating a chaotic scenario.



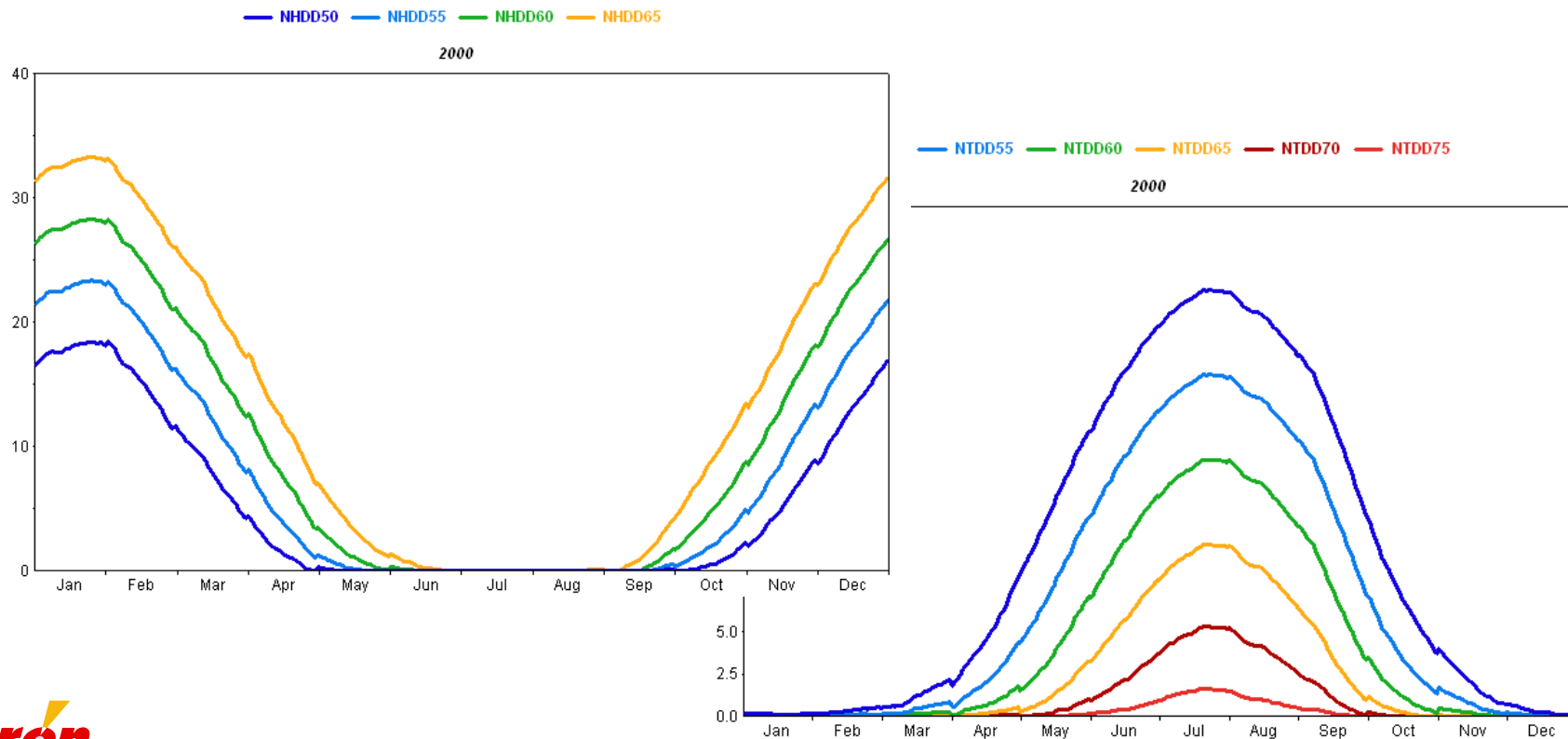
Averaging by Date

- Even after averaging across 30 years, average temperatures by date are a bit irregular. The same is true for HDD and CDD by date. Smoothing using centered moving averages will take the kinks out.



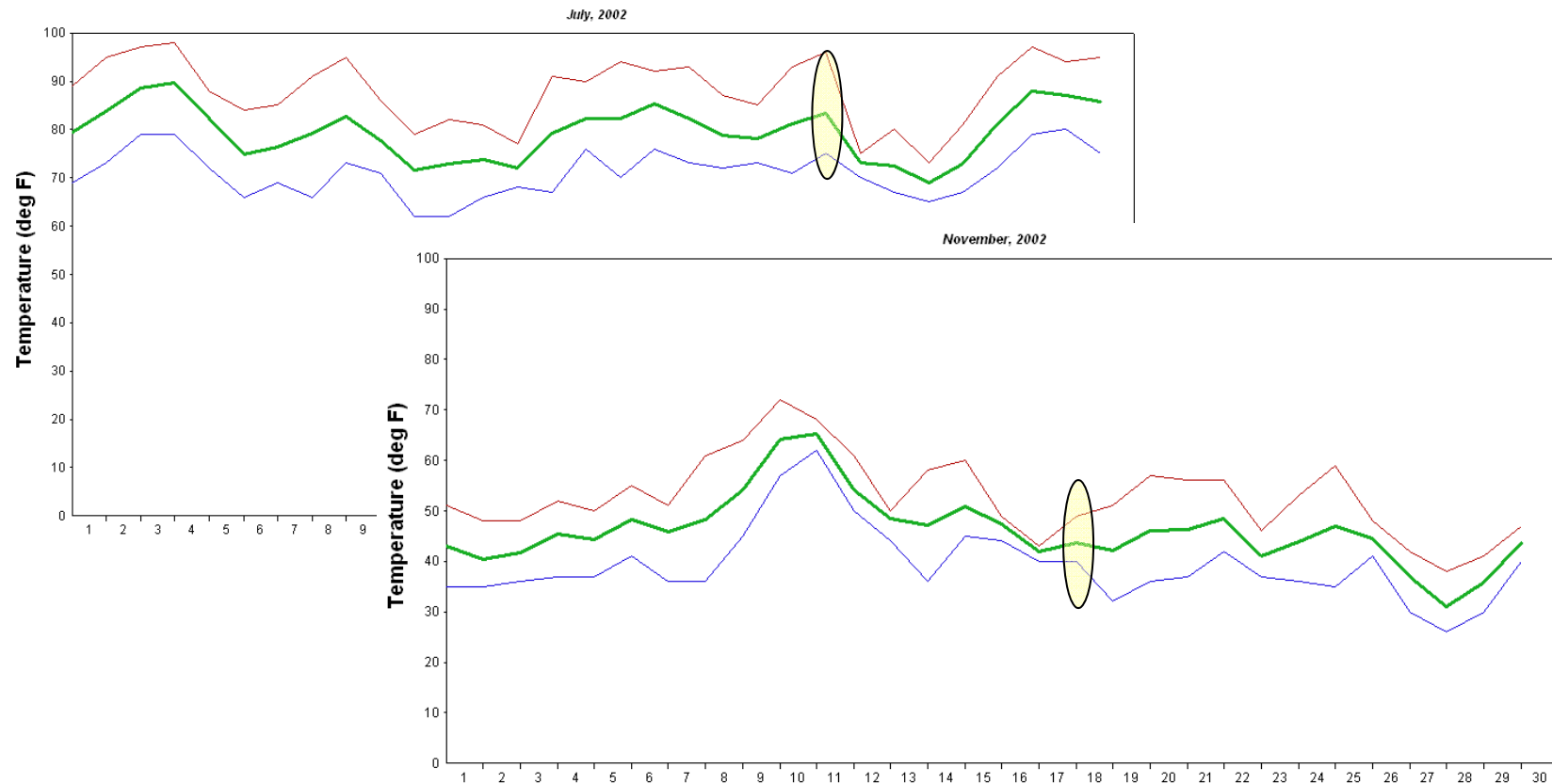
Smoothed Daily Normal HDD and CDD

- Smoothed daily normals look like this. By constructing these values for multiple CDD and HDD cut points, it is possible to distinguish between high powered and low powered degrees in the normalization process.

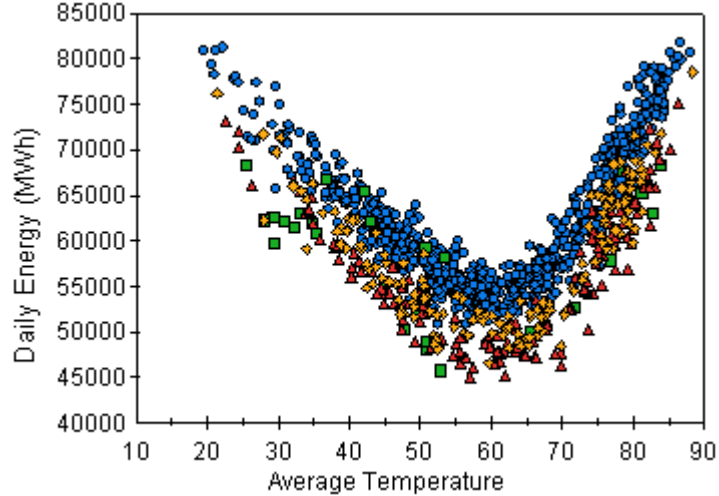


Peak producing weather is multidimensional

- Peak days are not necessarily the hottest or coldest days of the year. Peak producing weather is multidimensional – temperature, humidity, wind, clouds, prior day effects, ...



Why Weather Normalization is Tricky

- The response of energy usage to weather is not linear. As a result, averaging is dangerous. Accurate weather normalization requires nonlinear thinking.
 - Weather slopes are changing rapidly in some areas.
 - Weather response can differ significantly across months.
- 
- The scatter plot displays 'Daily Energy (MWh)' on the y-axis (ranging from 40,000 to 85,000) against 'Average Temperature' on the x-axis (ranging from 10 to 90). The data points form a U-shaped curve, with energy usage peaking at the extremes of temperature (around 20°F and 85°F) and reaching its minimum at moderate temperatures (around 60°F). The points are color-coded by month, with blue representing winter months (higher energy at low temperatures) and red/orange representing summer months (higher energy at high temperatures). This illustrates the non-linear and seasonal nature of energy usage in response to weather.
- The goal of weather normalization is to estimate the results of a simulation process with the following steps:
 - Estimate a model
 - Simulate energy use with the model for a variety of past weather patterns
 - Plot the distribution of energy outcomes
 - Compute the average of the distribution
 - Conceptually we should measure our methods against this goal

Overview of Itron Benchmarking Study

- Internet survey with about 20 questions
- 172 respondents
- 170 employ some type of weather normalization process. 2 did not employ any weather normalization process
- Distribution of respondents
 - 76% electric companies
 - 8% gas companies
 - 16% electric and gas companies

Q. What are weather normal sales used for?

- 170 responses
- Weather normalized sales are used by 73 percent of the respondents as the base for forecasting.
- 59 percent of respondents use normalization for variance analysis and 44 percent use normalization for financial reporting.
- Other uses for weather normalized sales were cited by 14 percent of respondents. These uses include:
 - support for rate filings;
 - class load studies; and
 - capacity planning.

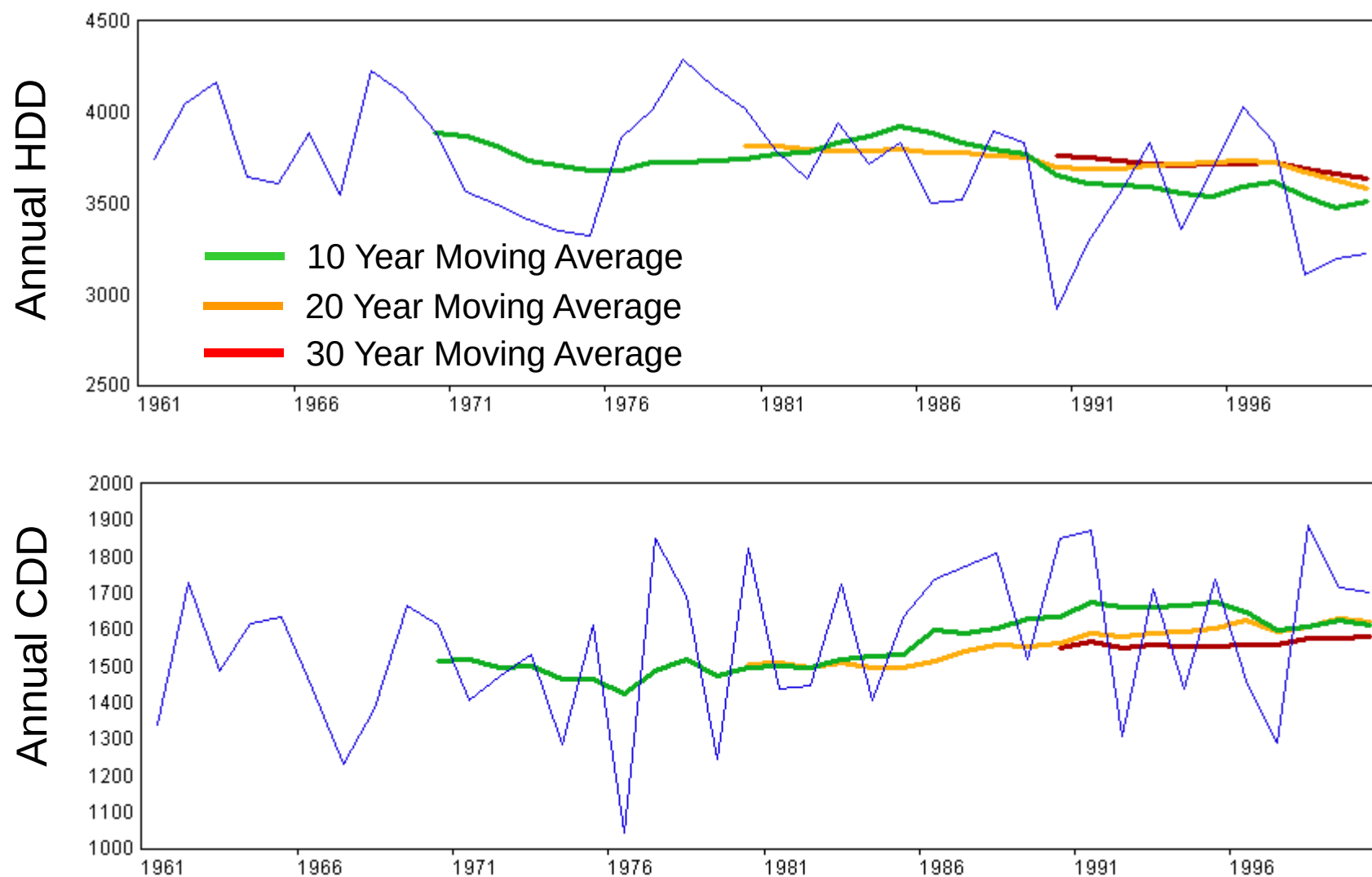
Q. What energy concepts do you weather normalize and report?

- The most common uses of weather normalization were reported to be:
 - Normalization of calendar month sales -- 59 percent
 - Normalization of system peak – 54 percent
 - Normalization of billed (cycle) sales – 45 percent
 - Normalization of unbilled sales – 13 percent

Definition of Normal Weather

- Alternative definitions of normal weather?
 - A representative weather pattern (e.g., Typical Meteorological Year – TMY)
 - Average or median values computed from past weather patterns.
 - Typical extreme values.
- Weather variables covered
 - Temperature, humidity, wind, indexes (THI, Wind chill, ...)
- Frequency of the normal weather variables
 - Hourly, daily, monthly, annual
 - Cycle vs. calendar
- Source of the normal weather results
- Number of years of data used to calculate normal values

Trends in HDD and CDD



Q. What do you use for normal weather?

- 167 responses
- Almost half (48 percent) of the respondents reported that they calculate normal weather using approaches developed by their company.
- The national weather services (NOAA and Environment Canada) are the source for roughly 40 percent of the respondents.
- Of this total, the vast majority use NOAA's 30-year normal weather.
- The remaining 11 percent of the respondents rely on commercial weather service providers or use typical year approaches.

Q. How many years of data do you use?

- 106 responses
- The majority (71 percent) of the respondents use 20 or more years to define normal weather.
 - Thirty years of data are used by 43 percent of the respondents.
 - Twenty years of data are used by 17 percent of the respondents.
 - Twenty-five years (2 percent) and
 - more than thirty years (9 percent) make up the balance of this majority.
- 6 percent of the respondents use 15 years of data.
- 16 percent use 10 years of data.
- Seven percent of the respondents use less than 10 years and most of those are using five years.
- The longest time span of data used is 60 years. The shortest span of data used is one year.

Q. Do you calculate an updated set of normal values every year?

- 114 responses
- Over two-thirds of the respondents (69 percent) update their normal values on a yearly basis
- The remaining 21 percent do not update their normal values annually

Q. Has the number of years that you use to define normal weather changed in the last few years?

- 115 responses
- Most respondents (75 percent) report that they have not changed their definition of normal weather.
- Of those that have changed the definition of normal years, the movement is, on average, to use a shorter data range.
- One of the reasons cited for change is that the 30-year normal is no longer relevant or does not reflect recent weather trends.
- Some cited the need to account for global warming as the reason behind moving toward a shorter time span.

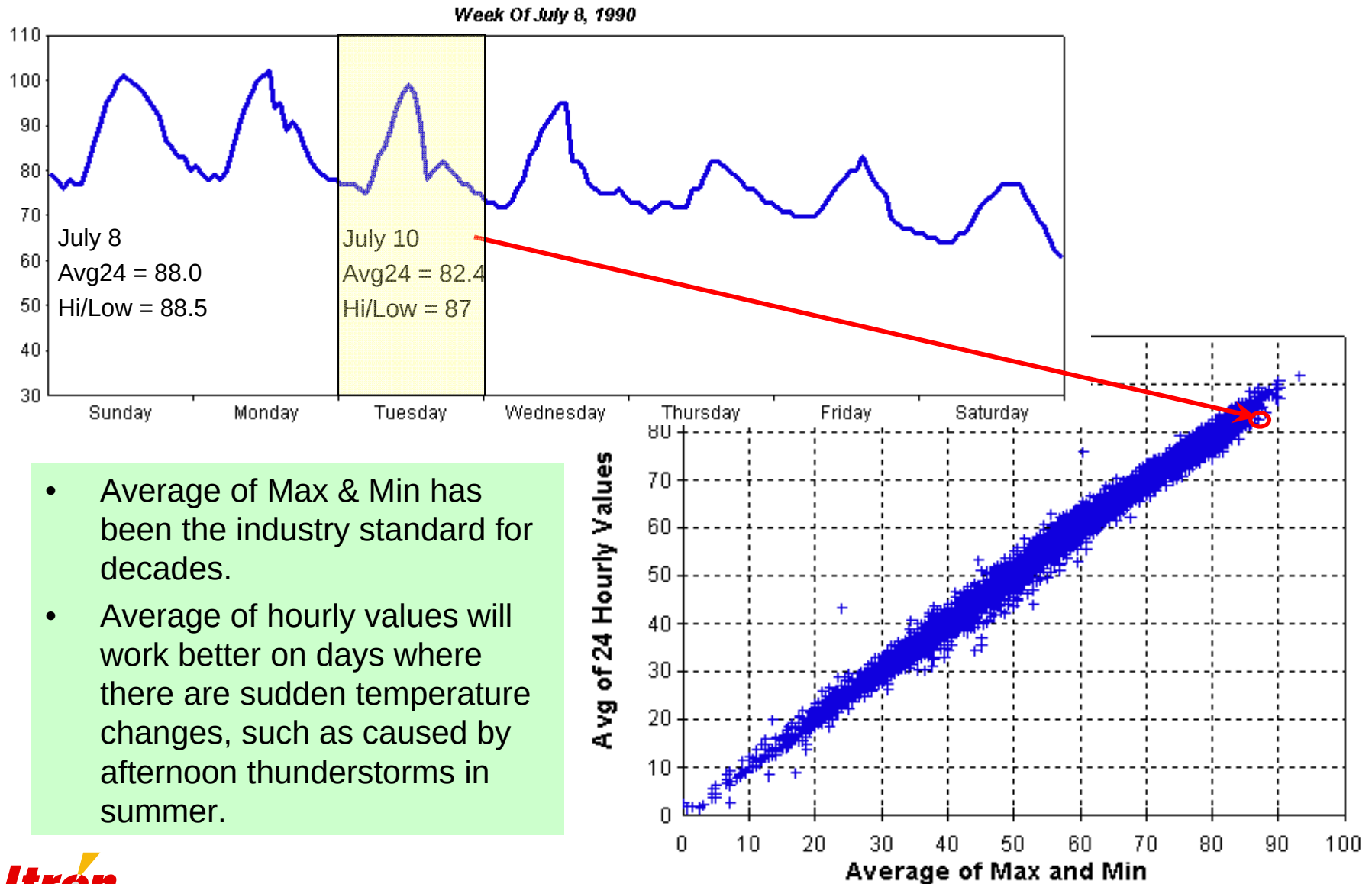
Definition of Weather Variables

- Heating Degree Day Variables
 - HD = heating degrees = degrees below a base temperature
 $HD = \text{Max}(55 - \text{AvgTemp}, 0)$
Example: AvgTemp = 43 implies HD = 12
 - HDD = heating degree days = Sum of HD over days in a cycle or month
- Cooling Degree Day Variables
 - CD = cooling degrees = degrees above a base temperature
 $CD = \text{Max}(\text{AvgTemp} - 65, 0)$
Example: AvgTemp = 77 implies CD = 12
 - CDD = cooling degree days = Sum of CD over days in a cycle or month
- Issues:
 - Definition of average temperature
 - Inclusion of other factors (humidity, wind)
 - Base temperature value(s)

Q. If you use HDD and/or CDD that are calculated using a daily average temperature, how do you calculate the daily average temperature?

- 137 responses
- The majority (60 percent) of respondents compute average daily temperature as the average of the daily maximum and minimum temperatures.
- The average daily temperature is computed using an average of the 24-hour temperature values by 19 percent of the respondents.
- Seventeen percent use the average daily temperature provided by their weather service provider.
- The remaining 4 percent of the respondents use some form of THI or weighting of multiple days.

Example of Averaging Methods



Calculation Logistics for Daily Normals

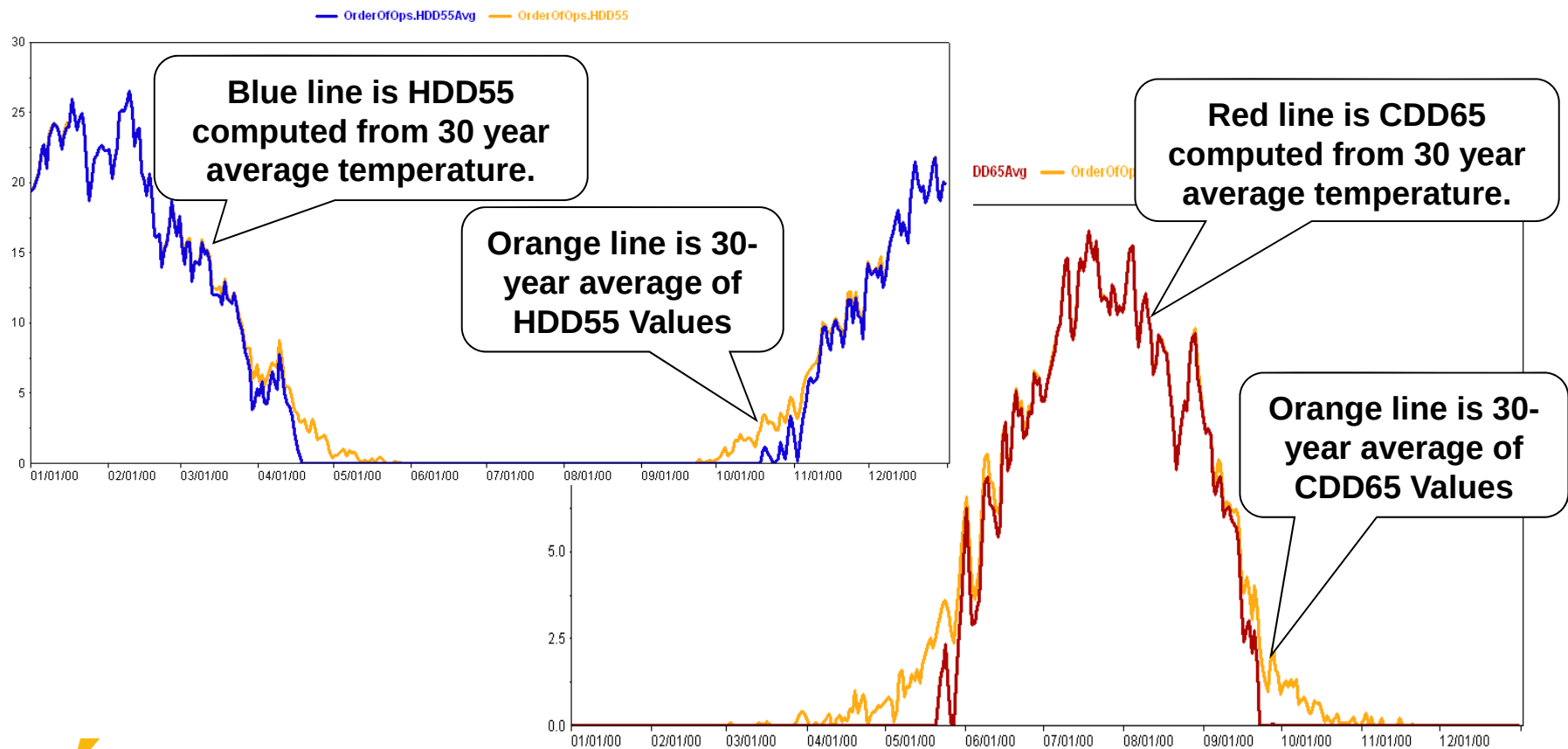
- Daily normals
 - NOAA has a daily normal series for HDD, CDD
 - Values are computed from 30 years and smoothed
 - DD values are computed for each date, then averaged across years
- The Rule: Always apply nonlinear operations first, then average. An example of the rule follows.

	Temp	CDD60	CDD65	CDD70
Year1	60	0	0	0
Year2	65	5	0	0
Year3	70	10	5	0
Year4	75	15	10	5
Year5	80	20	15	10
Average	70	10	6	3

- A generalization of the rule is that we should compute predicted load with each historical weather pattern (a nonlinear transformation of weather) and then average the predicted loads.

Order of Operations Matters

- The order of operations matters. Compute nonlinear transformations (HDD and CDD) first, then average over years. Computing HDD and CDD from averaged values will cause a downward bias in swing months.



Calculation Logistics for Monthly Normals

- Computed from Daily normals
 - Aggregate daily normal values by cycle
 - Average across cycles
 - Resulting normal monthly values reflect cycle schedules
- Computed from Monthly actuals
 - Compute DD variables for each historical month
 - Average across years for each month
 - Could also adjust for cycle days
- Calendar Month normals computed either way are about the same. The only fly in the ointment is Feb 29.
- Special calculations are required for peak producing weather.

Q. What variables do you use to adjust for cold weather?

- 166 responses
- The vast majority (80 percent) of the respondents report using Heating Degree Day (HDD) values based on temperature data.
- THI based HDD variables are used by 4 percent of the respondents.
- The remaining 16 percent of the respondents adjusted for temperature, humidity, cloud cover, and/or wind speed.

Q. If you use HDD, do you use a single base temperature or multiple segments?

- 86 responses
- The majority (85 percent) of the respondents report using a single trigger point.
- For companies using the Celsius scale, the average trigger point is 18°.
- The most commonly reported trigger points on a Fahrenheit scale are 55° and 65°.

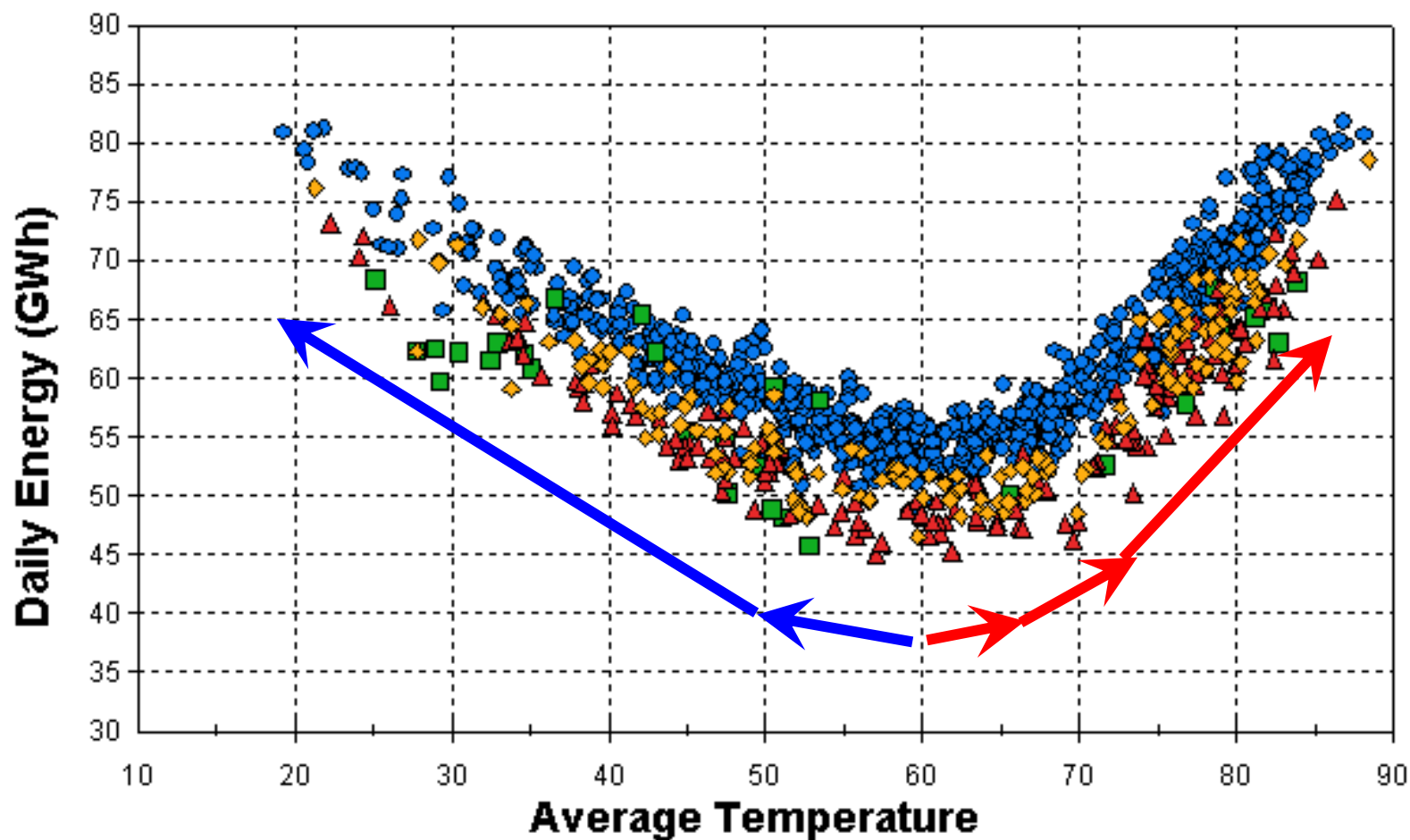
Q. What variables do you use to adjust for hot weather?

- 158 responses
- Roughly two-thirds (69 percent) of the respondents are using temperature data only to adjust for hot weather.
- Eight percent of the respondents use THI.
- The remaining 23 percent of the respondents report either that no adjustment is required for hot weather because they operate in climates with little to no cooling loads; or that they account for temperatures, humidity, cloud cover, and/or windspeed.

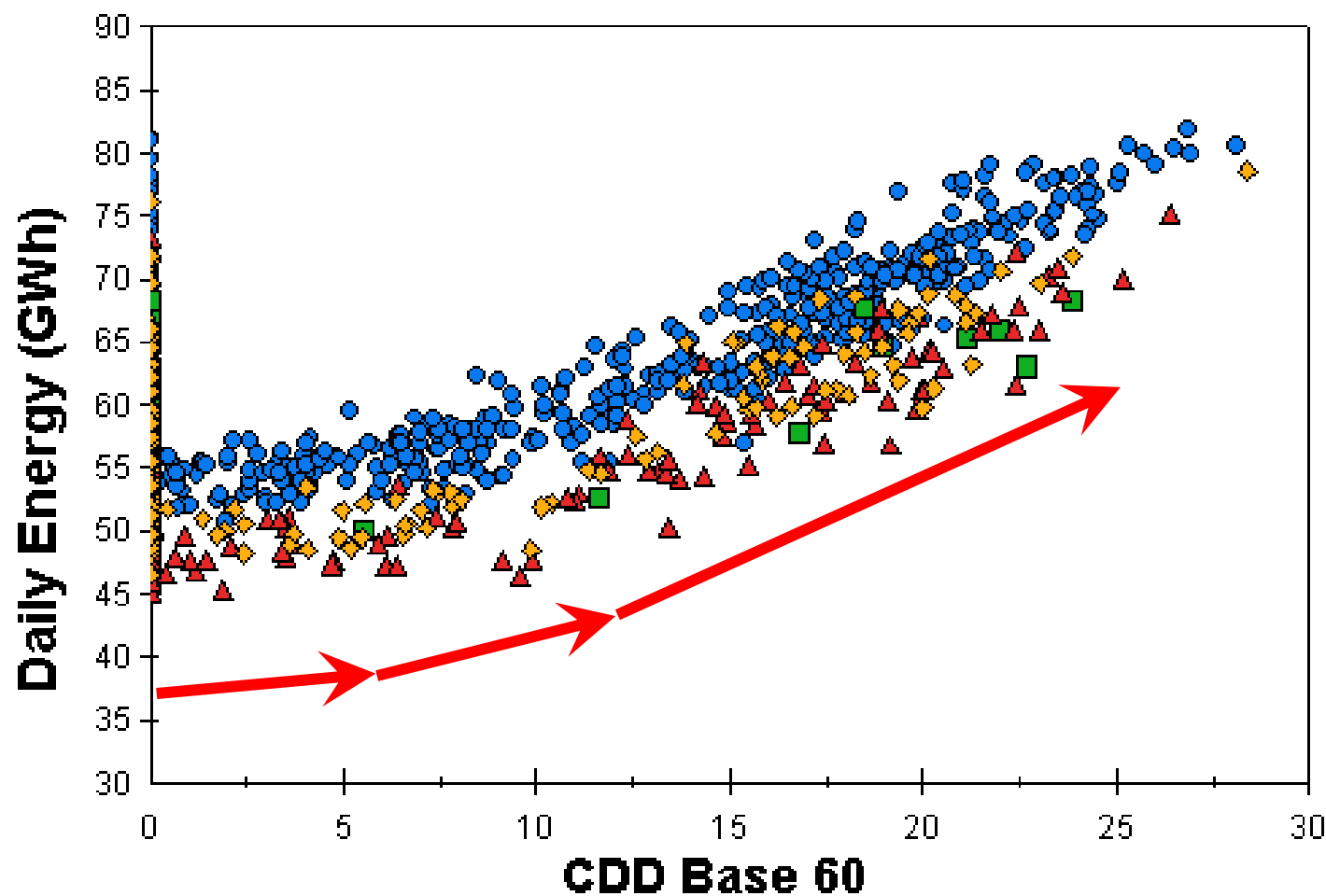
Q. If you use CDD, do you use a single trigger temperature or multiple segments?

- 123 responses
- Similar to the response for HDD, the vast majority (85 percent) use a single trigger point temperature for CDD.
- The typical trigger point is 18° Celsius or 65° Fahrenheit.
- For the respondents that use multiple trigger points, the trigger points that are used to separate cooling loads are between 65° F and 75° F and cooling loads above 75° F.

Example of Multiple Segments for HDD and CDD

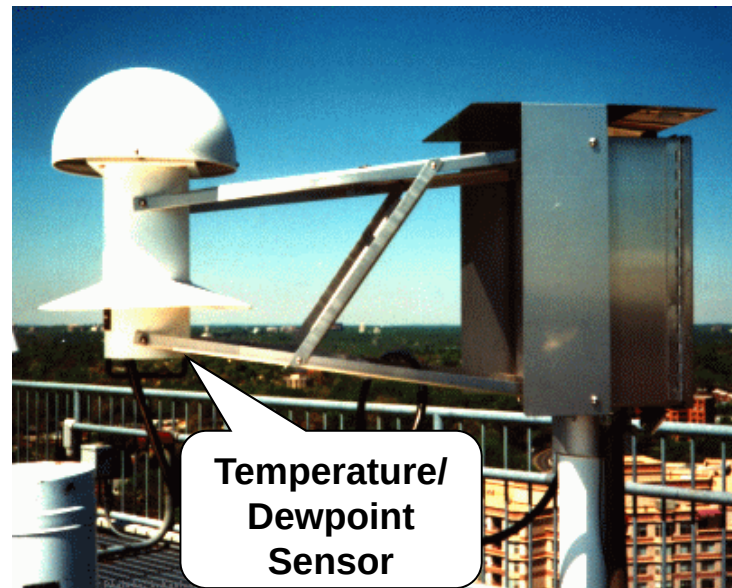


Example as a Regression Model Sees the Data



Q. If HDD and/or CDD are calculated using THI, which formula do you use to define THI?

- 108 responses
- (26 percent) $\text{THI} = 15 + 0.4 \times (\text{DryBulb} + \text{WetBulb})$
- (32 percent) $\text{THI} = 17.5 + 0.2 \times \text{DewPoint} + 0.55 \times \text{Drybulb}$
- (16 percent) $\text{THI} = \text{Drybulb} - (0.55 \times (1 - \text{Humidity})) \times (\text{Drybulb} - 58)$
- (26 percent) Other



Q. How do you compute weather normalized sales?

- 159 responses
- The majority (60 percent) use weather adjustment coefficients to normalize sales.
- Thirty percent use model simulations.
- The remaining 10 percent use some other process to normalize sales.

Model Approach

$$\text{Wthr Adj} = F(\text{Normal Weather}) - F(\text{Actual Weather})$$

Coefficient Approach

$$\text{Wthr Adj} = b * (\text{Normal Weather} - \text{Actual Weather})$$



Q. How often do you update your weather normalization coefficients or models?

- 163 responses
- The majority (52 percent) update coefficients once a year.
- Monthly updates are generated by 13 percent of the respondents and quarterly updates by 3 percent of the respondents.
- The remaining 31 percent of the respondents update coefficients every few years or in response to a rate case.

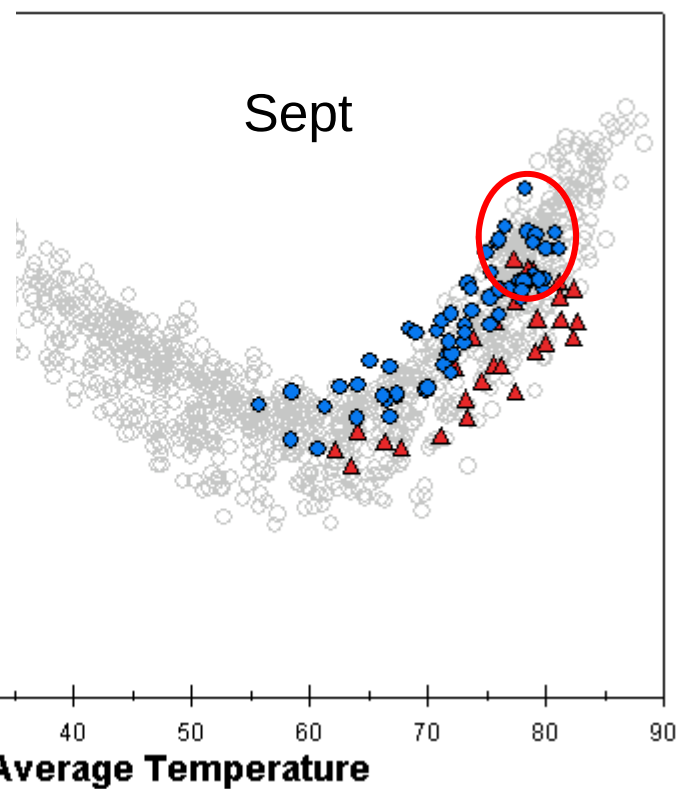
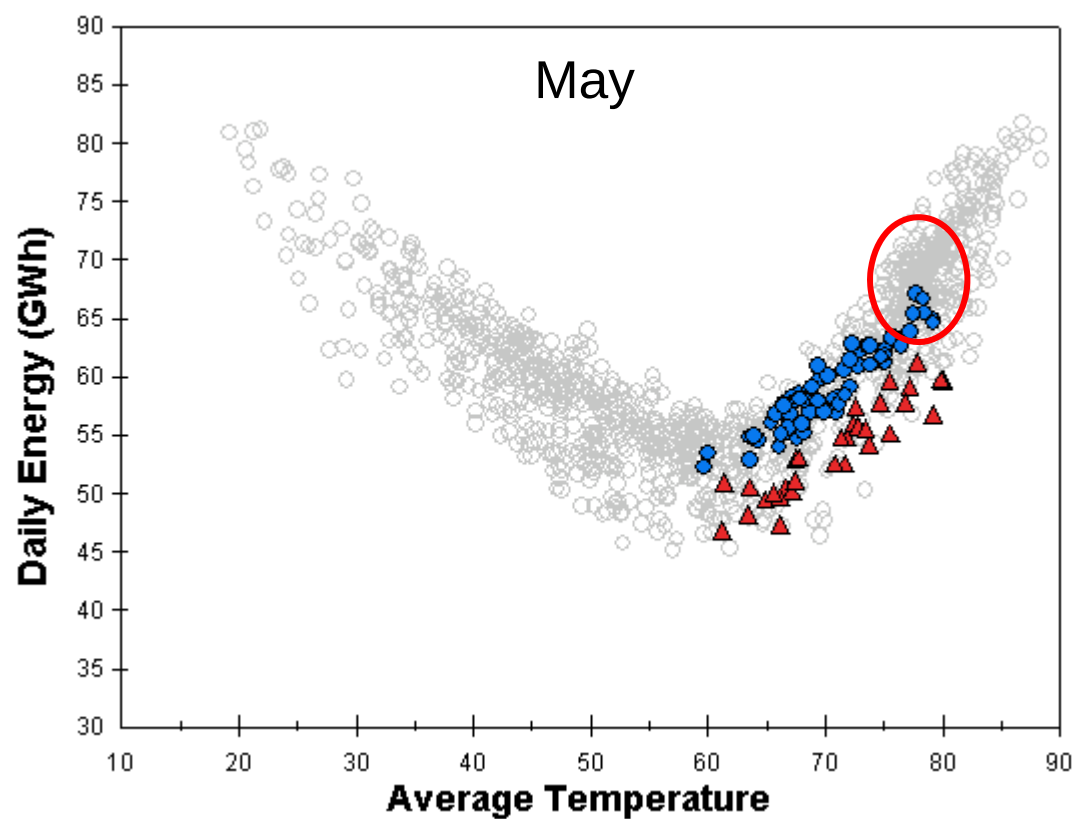
Q. Do you also re-calculate historical weather normal sales when you update your weather coefficients or models?

- 158 responses
- The majority (56 percent) of respondents do not re-calculate historical weather normal sales when they update their weather coefficients or models.
- Forty-four percent do re-calculate.

Q. Do weather adjustment parameters vary by month or season?

- 162 responses
- The majority of the respondents (66 percent) report that the weather adjustment parameters are allowed to vary either by month (45 percent) or by season (21 percent).
- Another 27 percent indicate that the parameters do not vary. Those that indicate some other method report that, for the most part, the parameters are allowed to vary although not necessarily by month or season.

Example of Weather Data by Month



Q. Is the weather normalization process that you use specified by your state commission?

- 166 responses
- Roughly 13 percent of respondents use a commission specified normalization process.
- Eighty-seven percent cited that they use their own weather normalization process.

Q. In statistical models of class sales (e.g., residential, commercial, industrial) which do you use?

- 158 responses
- Calendar-month HDD and CDD variables are used by 40 percent of the respondents.
- An additional 12 percent use calendar-month HDD and CDD variables with weights on current and previous months.
- Billing-cycle weighted HDD and CDD variables are used by 32 percent of the respondents.
- The remaining respondents use some form of lining up the HDD and CDD data with the sales data that is being modeled. Others indicated using the system load as a means of weighting the weather data.

Q. Is the approach you use for normalizing energy the same as what is used for developing normalized peaks?

- 159 responses
- The same approach was used to normalize energy and peaks by 42 percent of the respondents.
- The majority of the respondents (57 percent) indicate that they use different approaches.
- A small fraction of respondents indicate that they do not normalize peaks.

SUMMARY AND KEY ASSUMPTIONS

- ✧ Alberta is entering recovery mode. After contracting by an estimated 2.5% in 2009, its first recession since 1986, the Alberta economy is expected to expand by 2.6% in 2010. Average employment is forecast to increase by 15,000 in 2010; nonetheless, the unemployment rate is expected to remain at 6.6% because of offsetting growth in the labour force.
- ✧ Despite high oil prices and improved prospects for oil sands, the initial phases of Alberta's economic recovery are expected to be relatively moderate, in large part because of the subdued global economic recovery. Weak natural gas prices are also expected to weigh on investment in the conventional energy sector. Over the medium term, however, economic growth is forecast to strengthen to a sustainable pace of around 3% and the unemployment rate to fall below 5% by 2013.

Key Energy and Economic Assumptions

Fiscal Year Assumptions	2008-09 Actual	2009-10 Estimates	2010-11	2011-12	2012-13
Prices					
Crude Oil Price					
WTI (US\$/bbl)	85.94	69.95	78.75	83.50	89.50
Alberta Wellhead (Cdn\$/bbl) ^a	82.82	67.40	72.20	76.48	82.00
Bitumen @ Cold Lake (Cdn\$/bbl) ^b	61.54	49.36	56.38	59.79	64.07
Natural Gas Price					
Alberta Reference Price (Cdn\$/GJ)	6.97	3.40	4.25	5.25	5.50
Production					
Conventional crude oil (000s barrels/day)	497	454	424	419	419
Raw bitumen (000s barrels/day)	1,292	1,528	1,771	1,972	2,140
Natural gas (billions of cubic feet)	4,903	4,682	4,284	4,040	3,858
Interest rates					
3-month Canada treasury bills (per cent)	1.74	0.25	1.30	3.30	4.00
10-year Canada bonds (per cent)	3.40	3.45	4.05	4.70	5.00
Exchange Rate (US\$/Cdn\$)	89.59	91.67	95.00	95.00	95.00
Calendar Year Assumptions	2009 Actual	2010	2011	2012	2013
Gross Domestic Product					
Nominal (millions of dollars)	240,156 ^c	259,165	278,882	297,382	317,126
per cent change	-17.5	7.9	7.6	6.6	6.6
Real (millions of 2002 dollars)	181,114 ^c	185,875	191,301	196,980	203,069
per cent change	-2.5	2.6	2.9	3.0	3.1
Other Indicators					
Employment (thousands)	1,988	2,003	2,043	2,084	2,120
per cent change	-1.3	0.8	2.0	2.0	1.7
Unemployment rate (per cent)	6.6	6.6	6.0	5.3	4.9
Average Weekly Earnings (per cent change)	3.2 ^c	3.0	3.0	3.2	3.2
Personal Income (per cent change)	1.3 ^c	3.4	5.6	7.1	6.6
Corporate Profits (per cent change)	-58.2 ^c	34.1	17.7	6.8	8.6
Housing starts (number of units)	19,217 ^d	23,100	24,200	25,400	26,200
Alberta Consumer Price Index (per cent change)	-0.1	2.0	2.1	2.1	2.1
Population (thousands)	3,688	3,744	3,806	3,871	3,938
per cent change	2.6	1.5	1.7	1.7	1.7

^a Refers to the average price per barrel of Alberta light, medium and heavy oil.

^b The bitumen price is an estimate of the price at Cold Lake and is not an actual market price.

^c Alberta Finance and Enterprise estimate.

^d Preliminary actual.

HOUSING MARKET INFORMATION

NORTHERN HOUSING REPORT



CANADA MORTGAGE AND HOUSING CORPORATION

Date Released: 2012

As a component of its product suite, CMHC Market Analysis produces an annual Northern Housing Report. This report focuses on the new home construction market, existing home market, and rental markets in Whitehorse, Yellowknife, and Iqaluit.

Whitehorse

Economy

In 2011, real Gross Domestic Product (GDP) in the Yukon expanded by 5.6 per cent, following a four per cent increase in 2010. This represented the second fastest growing economy in Canada behind Nunavut, well ahead of the national growth rate of 2.6 per cent. This year, the pace of expansion in the Yukon economy will be more in line with the national average. While overall investment growth will moderate, higher levels of investment in key sectors such as mining and oil and gas will drive growth in 2012.

As was the case in 2011, the mining sector will continue to be a strong contributor to the Yukon's prosperity in 2012. Elevated metal prices, especially for gold, silver, and copper have encouraged expansion in the mining sector with investment levels expected to increase this year. Currently, there are three mines in operation with

more planned for the future. Statistics Canada's latest survey of Private and Public Investment Intentions indicates that total investment in the Yukon is expected to remain relatively consistent with 2011 levels. Higher investment across sectors such as mining, oil and gas extraction, housing, construction, and retail trade will more than offset declines in public administration and utilities.

Yukon employment grew in 2011, averaging a record high of over 19,000 people employed and 1,500 positions added. In the first quarter of 2012, the labour force posted gains while employment slightly decreased on a year-over-year basis. As a result, the unemployment rate in the first quarter of 2012 increased to 7.8 per cent, up from 4.5 per cent a year earlier. Moving forward, increasing investment in mining production and developments, as well as in construction, will contribute to the region's employment growth over the balance of the year.

Highlights

Whitehorse

- The elevated pace of residential construction in 2011 will continue in 2012.
- Existing home sales to moderate in 2012 yet remain elevated.
- Average vacancy rate is forecast to remain low at 1.5 per cent, pushing rental rates higher.

Yellowknife

- Housing starts will rise 13 per cent this year.
- Residential sales will remain on the growth trajectory.
- Despite rising vacancies, rental demand will remain strong this year, supporting higher rents.

Iqaluit

- The average two-bedroom rent for units surveyed increased from \$2,265 in 2010 to \$2,356 in 2011.
- The vacancy rate in non-social housing units increased to 1.4 per cent in 2011, up from 0.4 per cent in 2010.

SUBSCRIBE NOW!

Access CMHC's Market Analysis Centre publications quickly and conveniently on the Order Desk at www.cmhc.ca/housingmarketinformation.

View, print, download or subscribe to get market information e-mailed to you on the day it is released. CMHC's electronic suite of national standardized products is available for free.

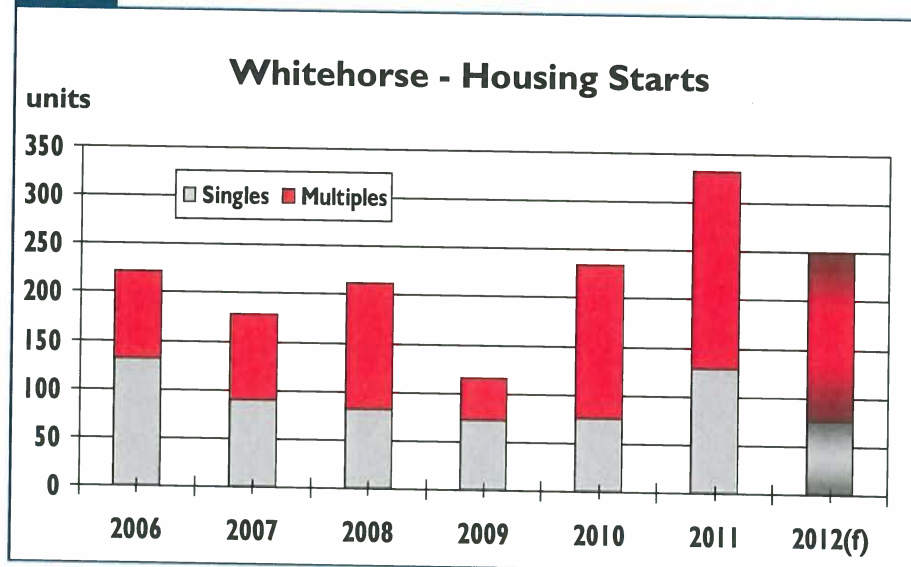
Positive net migration of over 300 people helped grow the Yukon's population in 2011 which reached 35,800 people, up 3.3 per cent from 34,667 the previous year. Most of the population growth in the Yukon came from a natural increase as births exceeded deaths. The Whitehorse population, which is 76 per cent of the total Yukon population, recorded positive growth of 3.4 per cent, totalling 27,190 persons at the end of 2011.

New Home Market

Total housing starts in Whitehorse reached a record in 2011 of 332 units, up 42 per cent from the previous year. Both single-detached and multi-family starts surpassed 2010 levels as new home production increased to meet elevated demand. Single-detached production increased 72 per cent over 2010 with 129 starts. Similarly, multi-family starts, which include semi-detached, row, and apartment units, increased 28 per cent over the previous year, reaching 203 starts. Row units accounted for the largest share of multi-family starts in 2011 at 93 units. Semi-detached and apartment units each totalled 48 and 62 units in 2011, up 14 and 46 starts, respectively.

The limited availability of serviced lots in Whitehorse will influence the level of construction in 2012. In 2010, the City of Whitehorse adopted a new Official Community Plan (OCP) which incorporates community development plans to address the elevated demand for housing. In order to accommodate household growth rates, the City of Whitehorse held land lotteries for infill sites to increase neighbourhood density. New residential development areas were also introduced, the largest being Whistle Bend, a six phase neighbourhood which can

Figure 1



Source: CMHC, CMHC Forecast (f)

hold an overall population of 8,000. However, lot sales for Phases 1 and 2 for Whistle Bend are not projected to occur until late 2012 and into 2013, with the first round of sales anticipated this fall. As such, starts this year are expected to take place on lots sold through previous lotteries, as well as infill sites and the first phase of Whistle Bend. In the first quarter of 2012, only four units have been started within the City compared to seven one year earlier. By year-end, we expect single-detached starts to total 75 units, a 42 per cent reduction from 2011.

The brisk pace of multi-family production from 2011 is expected to moderate in 2012 with rising supply levels and additions to inventory. However, under the sustainability directives of the 2010 OCP, the City of Whitehorse is actively engaged in the development of multi-family housing as an alternative to single-detached homes. Overall, multi-family starts are expected to remain elevated when compared to the five-year average of 125 units. With 32 multi-family starts in the first three

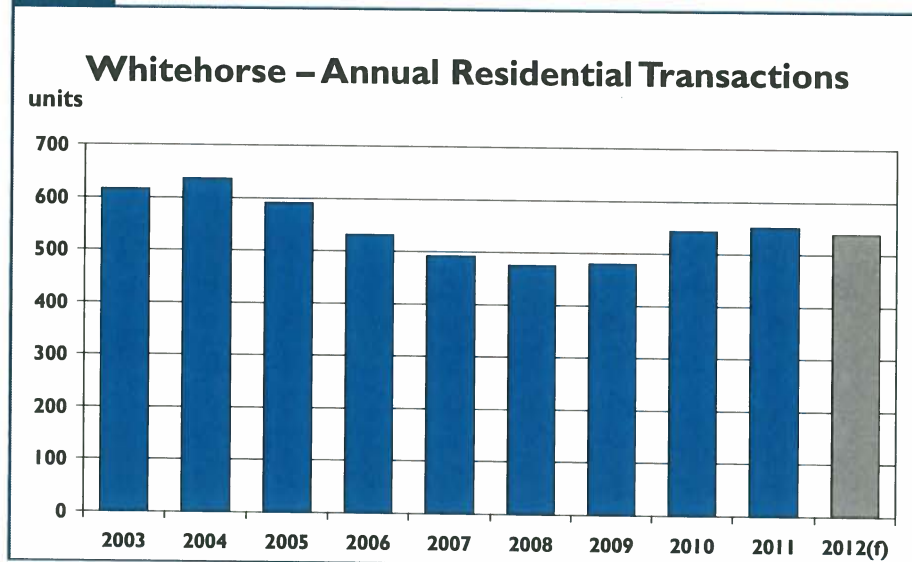
months of 2012, all in the form of row units, multi-family starts in 2012 are expected to total 175 units, down from 203 in 2011.

Real Estate Transactions

According to the Yukon Bureau of Statistics' quarterly Real Estate Survey, residential real estate transactions in Whitehorse, excluding those 'not-at-arm's-length', totalled 552 units in 2011, up 1.7 per cent year-over-year. Elevated housing demand along with limited lot availability brought additional buyers to the resale market. Both condominium and duplex sales in 2011 exceeded 2010 levels, with condominium sales increasing 16 per cent from 138 transactions in 2010 to a record high of 160 in 2011. Duplex transactions closed the year at 72 transactions, up 38.5 per cent from 52 units in the previous year.

Resale market transactions in 2012 are not expected to be as buoyant as the previous year. Despite elevated housing demand in Whitehorse, higher prices will temper resale market transactions. At the same

Figure 2



Source: Yukon Bureau of Statistics, CMHC Forecast (f)

time, the anticipation of increased new lot selection later this year will draw buyers back to the new home market. Total MLS® sales for the first quarter of 2012 were 14 per cent lower than the previous year at 67 transactions, compared to 78 in 2011. On the supply side, active listings exhibited large increases during the quarter, averaging 149 units per month compared to an average of 54 units the previous year. As such, resale transactions are anticipated to reach 540 units in 2012, a decrease of 2.2 per cent from 2011.

The average price for all types of housing in 2011 was \$362,058, up 11.8 per cent from 2010. Single-detached homes recorded a 15.8 per cent increase over 2010, averaging \$434,694. Meanwhile, the average prices of condominiums and duplexes posted year-over-year gains of 6.9 and 21.1 per cent, reaching record highs of \$270,850 and \$329,883, respectively.

After the 12 per cent gain in 2011, elevated levels of supply are expected to slow resale price growth this year. The average price of all types of

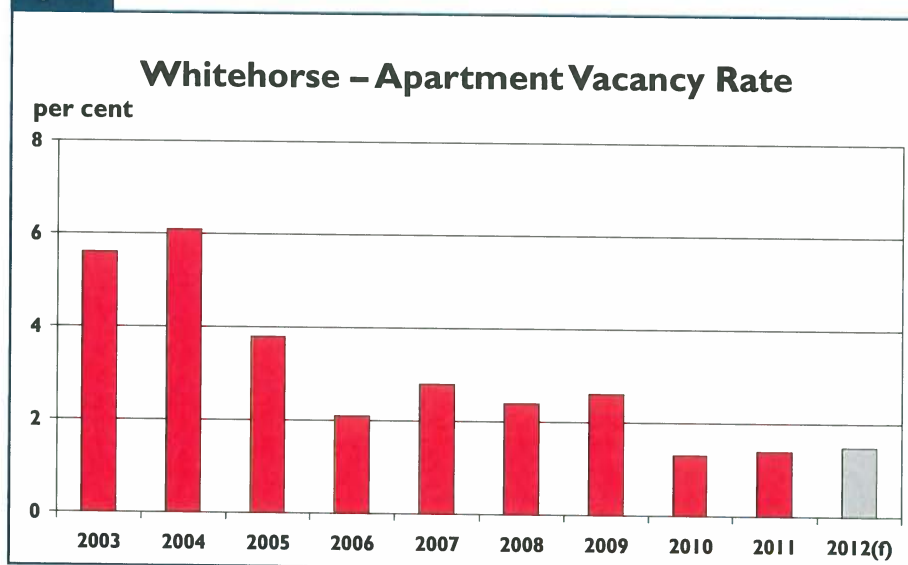
housing exhibited single-digit growth during the first quarter of 2012. Although the high level of supply relative to demand is impacting prices, the sale of new homes on the resale market is lifting the average price. Overall, Whitehorse's average resale price is forecast to increase five per cent to reach \$380,000 in 2012.

Rental Market

Prior to the fall of 2011, the Yukon Bureau of Statistics surveyed only buildings with four or more units. This has since changed to include coverage of any building with three or more rental units. As such, additional buildings were enumerated increasing the rental universe in the March 2012 survey. Based on this change, the figures released prior to 2012 are not strictly comparable to subsequent numbers.

Demand for rental units in Whitehorse continues to be elevated due to the high pace of economic activity. The average apartment vacancy rate in Whitehorse increased modestly in March 2012 to 1.3 per cent, compared to one per cent the previous year. This represents 13 vacant units out of 1,010 units in the apartment rental universe. CMHC does not expect any major changes to the rental universe for 2012 and as such, low vacancy rates are expected to prevail throughout the year with a vacancy rate of 1.5 per cent by December.

Figure 3



Source: Yukon Bureau of Statistics Q4 Yukon Rent Survey, CMHC Forecast (f)

After remaining stable throughout 2010 and exhibiting modest increases between \$10 and \$13 in 2011, the

median rent for a two-bedroom unit in Whitehorse increased to \$850 as of March 2012, up \$50 on a year-

over-year basis. CMHC forecasts the average two-bedroom rental rate will reach \$860 by December 2012.

Whitehorse Forecast Summary June 2012

	2009	2010	2011	% Change	2012(F)	% Change
New Home Market						
Total Starts (units)	116	234	332	41.9%	250	-24.7%
Single-Detached Starts	73	75	129	72.0%	75	-41.9%
Multi-Family Starts	43	159	203	27.7%	175	-13.8%
Real Estate Transactions*						
Residential sales	479	543	552	1.7%	540	-2.2%
Average price (\$)	288,386	323,918	362,058	11.8%	380,000	5.0%
Rental Market(Q4 data)**						
Apartment Vacancy Rate (%)	2.6%	1.3%	1.4%		1.5%	
Average Two-Bedroom Apartment Rent (\$)	800	800	813		860	

*MLS®+Private(source: YBS, CMHC Forecast)

** source: YBS, CMHC Forecast

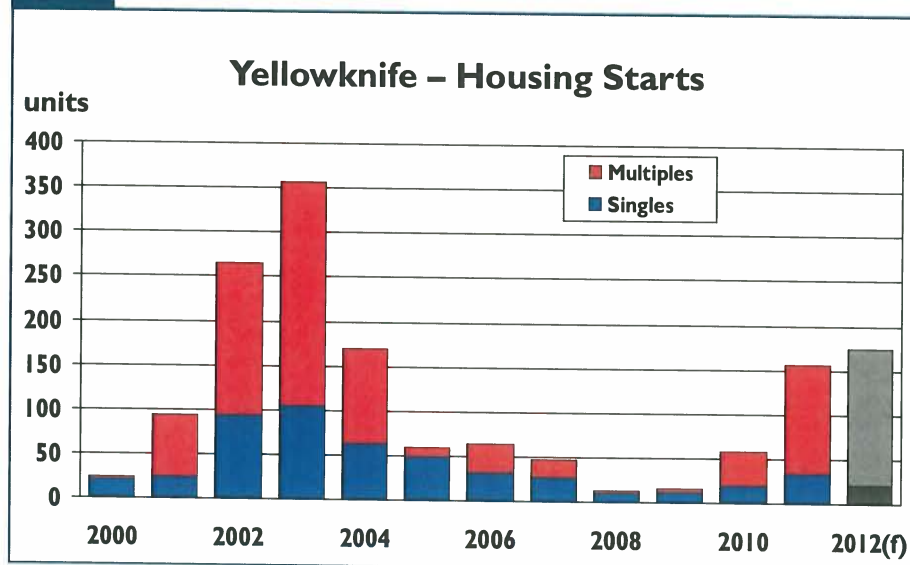
Yellowknife

Economy

Following a 1.3 per cent gain in real Gross Domestic Product (GDP) in 2010, the Northwest Territories' (NWT) economy slowed down in 2011. Mining output and oil and gas extraction decreased a combined total of 13 per cent year-over-year in 2011, while construction declined 5.3 per cent. This contributed to a 5.5 per cent reduction in NWT's GDP. The retreat in diamond output, which represents a substantial portion of mining activities, was chiefly due to an 18 per cent year-over-year decline in production from Ekati, the oldest of the NWT's three diamond mines. With diamond mining accounting for more than 25 per cent of NWT's GDP, 2011's decline in diamond production countered the 21 per cent gain in support activities for mining and oil and gas extraction, contributing to the overall economic contraction. In addition, last year's increase in construction activity at new mines was not strong enough to offset the decline in commercial and institutional construction.

A 17 per cent decrease in capital spending in 2011 also contributed to NWT's economic contraction. Public capital investment declined by 15 per cent while private capital expenditures decreased 18 per cent year-over-year. NWT's economy is expected to improve in 2012. The Conference Board of Canada forecasts real GDP growth of 5.9 per cent this year, spurred by rising diamond production and higher private capital investment. Gains in 2012 investment intentions confirm that NWT's economy is on track to grow. NWT's public and private organizations reported that they intend to invest \$1,293.2 billion in capital construction and machinery

Figure 1



Source: CMHC, CMHC Forecast (f)

and equipment in 2012, up 22 per cent from a year earlier.

Despite the moderation in economic output, NWT's economy created jobs in 2011. On average, 22,700 persons were employed in the territory in 2011 compared with 21,500 persons in 2010. The 1,200 new jobs added to the economy over this period translated into a 5.6 per cent year-over-year gain. A total of 1,300 people entered the labour force in 2011, raising the participation rate to 76.3 per cent from 71.8 per cent in 2010. Last year, additions to the labour force surpassed the number of jobs created by 100 persons, resulting in a negligible change to the unemployment rate. With the number of unemployed people up by only 100 persons, the jobless rate of 7.3 per cent in 2011 was unchanged from 2010.

Job creation is again expected for 2012. The NWT has reported that 200 new jobs were created between January 2011 and January 2012. Much of the impetus in housing demand in Yellowknife is supported by job creation and population growth. Early

data pointing to job growth thus far in 2012 confirms that job gains will continue to support housing demand. The 2.9 per cent increase in Yellowknife's population between 2006 (18,700 persons) and 2011 (19,234 people) is also underpinning housing demand.

One potential downside risk to housing demand in NWT remains out-migration. NWT continued to experience net out-migration last year with nearly 800 people leaving the territory on a net basis. With that, NWT's population declined from 43,830 persons in 2010 to 43,675 people in 2011. However, Yellowknife's population continued to grow over the last couple of years, indicating emigrants were mostly from other centres in NWT.

New Home Market

Gains in employment, the high level of incomes, and stability in mortgage carrying costs which had propelled residential construction in Yellowknife in 2010 continued in 2011. Following 19 single-detached starts in 2010,

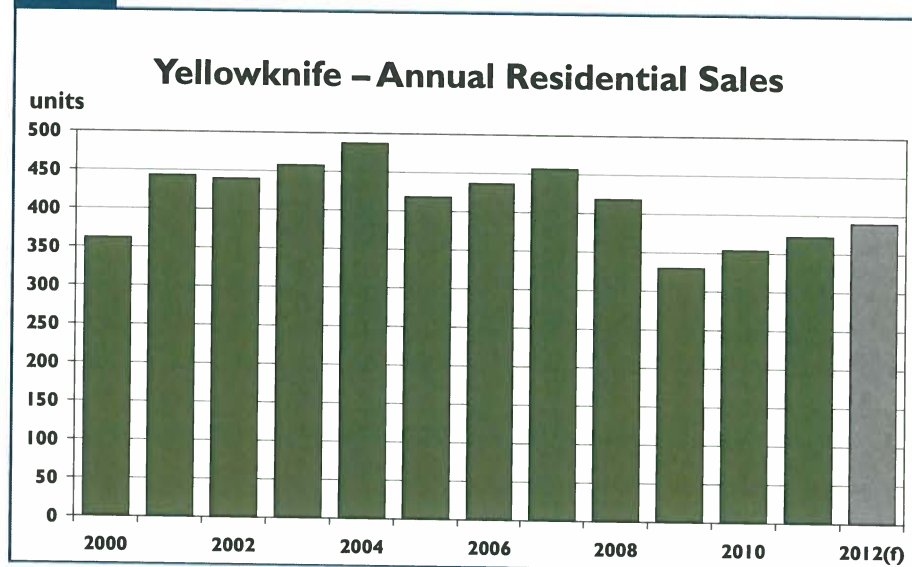
32 new single-detached homes broke ground in Yellowknife in 2011, representing a 68 per cent gain. Meanwhile, after a 660 per cent rise in 2010 to 38 homes, multi-family starts more than tripled in 2011 to 123 units. In total, the 155 new homes that were started in Yellowknife in 2011 represented a 172 per cent gain over 2010 and the strongest annual performance in seven years.

Continued job creation will provide additional support to the market this year and create the opportunity for a 13 per cent increase in residential construction. Additional lot selection, stable carrying costs, and elevated incomes will also serve to raise 2012's total housing starts to 175 units. Multi-family construction, which includes semi-detached, row, and apartment units, is set to grow this year, with already 27 multi-family starts through March. With the market becoming more demand-driven, multi-family builders will manage their 150 plus orders to spread work evenly throughout the year. With demand on a firm footing, builders will start 155 multi-family homes this year, a 26 per cent gain over 2011. Meanwhile, with five single-detached starts to the end of March, construction of single-detached homes is on track to reach the 20-unit mark this year, a 38 per cent reduction from 2011. The expected decline in single-detached starts this year should not be seen as a sign of weaker demand but rather a shift toward lower priced multi-family products.

Resale Home Market

Following a five per cent gain in 2011 to 372 homes, residential sales in Yellowknife will remain on the growth trajectory this year. Resales after the first quarter had experienced triple-digit year-over-year gains in

Figure 2



Source: CMHC (MLS® & Private Sales), CMHC Forecast (f)

2012. An improving job market will maintain a favourable climate for resales. CMHC is forecasting that 390 resale transactions will take place in 2012, a five per cent advance over 2011. Also supporting the upcoming expansion are stable carrying costs due to low mortgage rates and modest price growth, as well as a relatively low cost differential between home ownership and rental housing in Yellowknife. The difference between the monthly mortgage payment for an average resale home and the monthly average two-bedroom apartment rent was \$323 in 2011. With the cost difference relatively low, this will help provide some renters the opportunity to move into homeownership.

Higher job growth and demand for housing contributed to a steady turnover in stocks last year, which improved the demand/supply balance, allowing for some price gains. The outlook calls for this trend to persist throughout the forecast period. During the first quarter of 2012, the sales-to-new listings ratio (SNLR) was 177 per cent, up from 93 per cent a year earlier. Continued progress in the

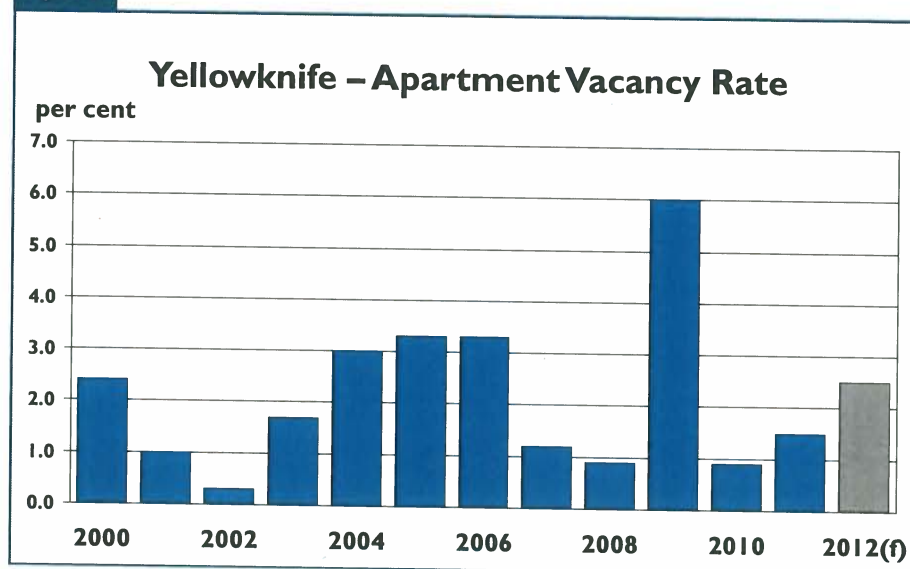
demand/supply balance will support further price growth in 2012. After a two per cent increase in the previous year to \$375,050, the average sale price is expected to rise by another two per cent in 2012 to \$382,900. While a slight reduction in listings and higher demand supported price growth last year, the overall increase was subdued due to a modest compositional shift away from the highest price range. In 2011, 32 per cent of sales were above \$400,000, compared to 2010 when the share was 34 per cent. The expected increase in prices this year will mostly be due to on-going strength in demand relative to supply levels. During the first quarter of 2012, 115 resale transactions took place, more than double the 53 resales this time last year. Meanwhile, new listings rose from 57 homes during 2011's first quarter to 65 units during the first quarter of this year. The composition will also inhibit average price growth this year. To the end March 2012, 51 per cent of sales were below \$300,000 compared to a 32 per cent share one year earlier.

Rental Market

Higher levels of net out-migration combined with the movement into homeownership have impacted rental demand in Yellowknife. The apartment vacancy rate in Yellowknife increased from 0.9 per cent in October 2010 to 1.5 per cent in October 2011. Both supply and demand factors contributed to the increase. On the supply side, 17 units were removed from the apartment rental stock between the October 2010 and 2011 surveys. On the demand side, the 12 per cent gain in sales of existing homes through September 2011 has coincided with some renters vacating their rental units. In addition, the net number of persons who left the Territories in the first half of 2011 was considerably higher year-over-year, and this has also contributed to the increase in vacancies. On average, two-bedroom apartments rented for \$1,566 per month in October 2011, up from \$1,486 a year earlier.

Rental demand will remain strong this year supported by job gains and population growth. However, rental

Figure 3



Source: CMHC, CMHC Forecast (f)

demand will not be strong enough to offset vacancies stemming from the movement into homeownership. With relatively stable mortgage carrying costs this year, many renters will move forward and buy an existing or new home or condominium. With rental units being freed as tenants move out, the average vacancy rate will rise to 2.5 per cent in October

2012 from 1.5 per cent in October 2011. Despite this increase, overall vacancies will remain relatively low. As such, landlords will likely consider increasing rents. Average rents of two-bedroom apartments will rise to \$1,610 in October 2012 from \$1,566 in October 2011.

Yellowknife Forecast Summary June 2012

	2009	2010	2011	% Change	2012(F)	% Change
New Home Market						
Total Starts (units)	16	57	155	171.9%	175	12.9%
Single-Detached Starts	11	19	32	68.4%	20	-37.5%
Multi-Family Starts	5	38	123	223.7%	155	26.0%
Resale Market*						
Residential sales	330	354	372	5.1%	390	4.8%
Res. resale price (\$)	326,600	367,250	375,050	2.1%	382,900	2.1%
Rental Market (October Survey)**						
Apartment Vacancy Rate (%)	6.0%	0.9%	1.5%		2.5%	
Average Two-Bedroom Apartment Rent (\$)	1,473	\$1,486	\$1,566		\$1,610	

*MLS®+Private(source CMHC Forecast)

** source: CMHC, CMHC Forecast

Iqaluit

Economy

Nunavut's economy continued to grow in 2011, thanks largely to continued expansion in the resource sector and strong metal prices. Following an 11.3 per cent expansion in 2010, real Gross Domestic Product (GDP) in Nunavut advanced by 7.7 per cent in 2011. Nunavut's mining sector increased substantially in 2011 with the opening of Agnico-Eagles Meadowbank gold mine, which produced over 300,000 ounces of gold. Meanwhile, several mining projects in various stages of development promoted exploration spending of nearly \$400 million in 2011, the second highest year for exploration spending in the territory's history. Spending intentions for mining in 2012 are estimated at \$568.6 million. This represents the fourth highest spending intention in the country and a 48 per cent increase over the previous year. The mining exploration spanned across many different types of natural resources such as iron ore, uranium, and nickel. It is expected that the mining industry will continue to benefit from growing global demand and be a strong contributing factor in Nunavut's economic growth this year and beyond. Real GDP is expected to

expand by more than 10 per cent in 2012.

Nunavut's strong economic expansion in 2011 contributed to solid gains in consumer spending. For the entire year, retail sales increased six per cent from 2010 levels. Wage growth continued in 2011, further supporting the increase in spending. Average weekly earnings for all industries increased 2.3 per cent year-over-year, from \$828 in 2010 to \$847 in 2011. Last year's growth in earnings was supported by a higher proportion of full-time job gains. As was the case in 2010, the goods-producing sector recorded the largest gain in wages last year, up 12 per cent year-over-year. The public sector remains a prominent fixture in the Nunavut economy, with average weekly earnings for public administration gaining two per cent in 2011.

Residential investment was another large contributor to economic growth in 2011. Investment in single-detached units, semis, rows, apartments, conversions, and renovations, increased 51 per cent to \$112 million in 2011 from \$74 million in 2010. Semi-detached residential investment experienced the largest increase, with spending tripling 2010 levels. Public and private investment intentions as surveyed by Statistics Canada indicate that total investment was estimated

to grow by 26 per cent in 2011, with a similar level of activity in 2012. Capital expenditures in the housing sector are expected to rise by nine per cent in 2012 over 2011.

Nunavut's population continued to grow in 2011, increasing over one per cent from 2010 levels. This was attributed to a stronger rate of natural increase (births minus deaths) than previous years. Meanwhile, following a modest gain in 2010, net migration declined in 2011. After adding 72 people in 2010, the territory witnessed negative net migration with 184 total net migrants leaving in 2011, mostly attributed to movement to other provinces. Modest employment gains likely weakened net migration last year. Across all industries, total employment in Nunavut rose one per cent on a year-over-year basis in 2011. Given a higher pace of economic growth in 2012, we can expect a stronger expansion of payrolls as we move forward.

New Home Construction

The construction of new homes in Iqaluit has been strong over the last several years, averaging roughly 170 permits a year since 2006. In 2011, the industry continued to be a large contributor to the region's economic performance. More than half of the lots in Phase 3 of the Plateau

Iqaluit Residential Transactions

	2008	2009	2010	2011
Residential Sales	52	37	43	60
Average Price	\$348,544	\$349,168	\$344,008	\$335,456
Median Price	\$355,000	\$356,350	\$369,000	\$367,000
Avg. Price Sq. foot	\$231.76	\$220.00	\$204.11	\$237.00

Source: CMHC

Subdivision began construction in 2011. Completion of these units is expected in 2012. In the Road to Nowhere subdivision, a major landlord has recently completed four eight-plexes. The intent of these new units was to attract higher-income occupants and the rents were priced accordingly. Building 4100 in the Road to Nowhere subdivision was completely destroyed by fire in March. It is now in the process of being replaced and expanded from 29 to 39 units. There is another apartment complex being developed adjacent to Building 4100, which will also be comprised of 39 units and have a similar configuration. Both of these buildings are expected to be complete by the summer of 2012.

The value of residential development permits issued in 2011 totalled \$68.8 million, more than double the amount of 2010. Permits were issued for 18 single-detached and 295 multi-family units in 2011. This compares to 21 single-detached and 52 multi-family permits in 2010.

Iqaluit Residential Transactions		
2011	Single-detached	Condo/Row/Semi
Residential Sales	37	23
Average Price	\$370,527	\$279,038
Median Price	\$370,000	\$229,000
Ave. Price Sq. foot	\$249.00	\$216.00
Source: CMHC		

Residential Transactions

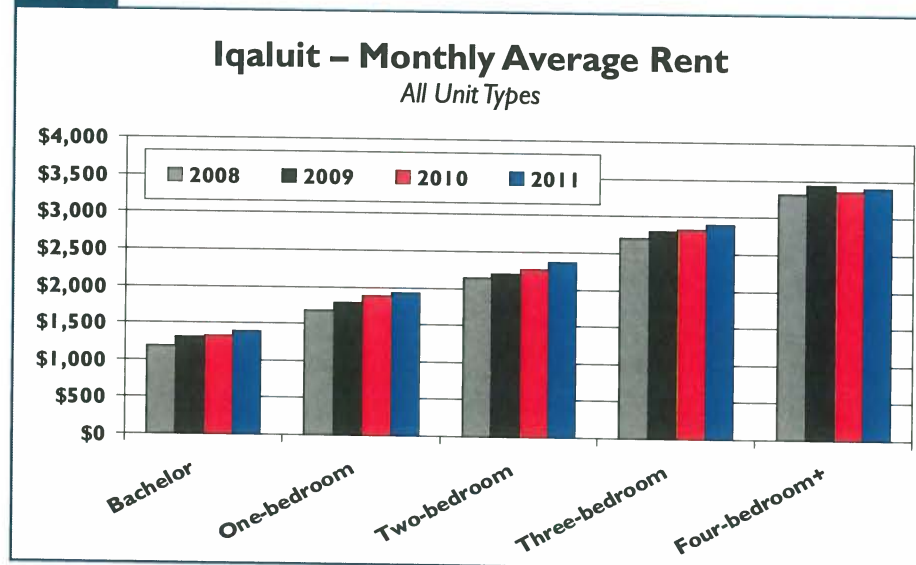
The continued economic expansion and employment gains supported demand for existing homes within the City of Iqaluit. In 2011, transfers of residential real estate in Iqaluit totalled 82 units, up significantly from the 48 transfers a year earlier. The Iqaluit housing market is unique compared to most markets across Canada. Like many Nunavut communities, most of the land or lots in Iqaluit are owned either by the municipality or by the Inuit. The majority of lots are titled to the Municipal Corporation which issues

Leasehold Titles for use. When the title of the lease is transferred, a price (either the cost of development or some consideration) is also determined.

In 2011, 22 of the 82 transfers were declared at nominal transactional value of between \$1 and \$10. The remaining 60 transactions declared a sale value between \$74,000 and \$490,000. Excluding transactions priced below \$10 to remove any compositional impact, the resulting average price of residential transactions was \$335,456 in 2011, down 2.5 per cent from the average in 2010. The reduction can be attributed to a shift in the sales mix to smaller homes and a lower proportion of single-detached sales which pushed down the average. The median price also decreased slightly from the previous year. In 2011, the median price of all transactions was \$367,000, down from \$369,000 in 2010.

The 82 residential transactions in 2011 were comprised of 59 single-detached and 23 multi-family units. Of the 59 single-detached homes, 22 reported a nominal transactional value of between \$1 and \$10. The remaining 37 recorded a sale value between \$74,000 and \$490,000. The average price for units with a sale price above \$74,000 was \$370,527, while the median price was \$370,000. For the

Figure 1



Source: CMHC Survey (does not include public housing units)

23 multi-family transactions, which were comprised of semi-detached, row and condominium units, sale prices ranged from \$169,000 and \$398,000, resulting in an average price of \$279,038. The median price of multi-family residential sales in 2011 was \$229,000.

Rental Market

CMHC's annual survey in Iqaluit revealed the rental universe increased from 1,793 units in 2010 to 1,809 units in 2011. As was the case in 2010, the additions to Iqaluit's rental universe resulted from the completion of newly constructed structures. Of Iqaluit's 1,809 rental units surveyed, 362 were allocated as social housing (public) units. While identified, these units were not included in any average rent calculations.

Following three consecutive years of decline, Iqaluit's rental vacancy rate increased slightly in 2011. Within structures surveyed by CMHC, the vacancy rate in Iqaluit's non-social housing units increased to 1.4 per cent in 2011, up from 0.4 per cent in 2010 and 0.6 per cent in 2009. Of the 1,447 non-social housing units surveyed, 20 were found vacant at the time of the 2011 survey.

Although the vacancy rate increased in 2011, it remained low enough to support continued increases in average rent. The 2011 survey revealed an increase in the average rent for all bedroom types, moving from \$2,268 in 2010 to \$2,324 in 2011. At \$56, the gain in average rent

Average Monthly Rent - by bedroom type (excluding social housing)			
	2010	2011	Yr/Yr Change
Bachelors	\$1,330	\$1,386	\$56
One-bedroom	\$1,889	\$1,930	\$41
Two-bedroom	\$2,265	\$2,356	\$91
Three-bedroom	\$2,820	\$2,891	\$71
Four-bedroom	\$3,357	\$3,402	\$45
Total Average	\$2,268	\$2,324	\$56

Source: CMHC

was higher than the previous year. In addition to low vacancies, several new buildings were constructed in 2011 resulting in higher rents. Furthermore, where long-term leases did not exist, landlords were able to increase rents to account for rising costs.

The strongest gain in average rent was among two-bedroom suites where rents increased 4.0 per cent from \$2,265 per month in 2010 to \$2,356 in 2011. The lowest increase occurred in four-bedroom suites where the average rent increased marginally from \$3,357 in 2010 to \$3,402 in 2011.

Rental Market Survey Methodology

The purpose of the CMHC survey was to identify residential accommodation in Iqaluit available for long-term rental and identify the current market rent on those units. The survey was conducted over a

Rental Universe - 2011	
Units	
Bachelors	63
One-bedroom	495
Two-bedroom	493
Three-bedroom	360
Four-bedroom	36
No BR Identified (Social)	362
Total	1,809

Source: CMHC

six-week period, from mid-November to year-end 2011. Our survey depends on the input of developers, building owners, two senior levels of government (Territorial and Federal) and their respective housing officials. CMHC acknowledges their hard work and providing timely and accurate information.

CMHC—HOME TO CANADIANS

Canada Mortgage and Housing Corporation (CMHC) has been Canada's national housing agency for more than 65 years.

Together with other housing stakeholders, we help ensure that the Canadian housing system remains one of the best in the world. We are committed to helping Canadians access a wide choice of quality, environmentally sustainable and affordable housing solutions that will continue to create vibrant and healthy communities and cities across the country.

For more information, visit our website at www.cmhc.ca

You can also reach us by phone at 1-800-668-2642 or by fax at 1-800-245-9274.
Outside Canada call 613-748-2003 or fax to 613-748-2016.

Canada Mortgage and Housing Corporation supports the Government of Canada policy on access to information for people with disabilities. If you wish to obtain this publication in alternative formats, call 1-800-668-2642.

The Market Analysis Centre's (MAC) electronic suite of national standardized products is available for free on CMHC's website. You can view, print, download or subscribe to future editions and get market information e-mailed automatically to you the same day it is released. It's quick and convenient! Go to www.cmhc.ca/housingmarketinformation

For more information on MAC and the wealth of housing market information available to you, visit us today at www.cmhc.ca/housingmarketinformation

To subscribe to priced, printed editions of MAC publications, call 1-800-668-2642.

©2012 Canada Mortgage and Housing Corporation. All rights reserved. CMHC grants reasonable rights of use of this publication's content solely for personal, corporate or public policy research, and educational purposes. This permission consists of the right to use the content for general reference purposes in written analyses and in the reporting of results, conclusions, and forecasts including the citation of limited amounts of supporting data extracted from this publication. Reasonable and limited rights of use are also permitted in commercial publications subject to the above criteria, and CMHC's right to request that such use be discontinued for any reason.

Any use of the publication's content must include the source of the information, including statistical data, acknowledged as follows:

Source: CMHC (or "Adapted from CMHC," if appropriate), name of product, year and date of publication issue.

Other than as outlined above, the content of the publication cannot be reproduced or transmitted to any person or, if acquired by an organization, to users outside the organization. Placing the publication, in whole or part, on a website accessible to the public or on any website accessible to persons not directly employed by the organization is not permitted. To use the content of any CMHC Market Analysis publication for any purpose other than the general reference purposes set out above or to request permission to reproduce large portions of, or entire CMHC Market Analysis publications, please contact: the Canadian Housing Information Centre (CHIC) at chic@cmhc.ca; 613-748-2367 or 1-800-668-2642.

For permission, please provide CHIC with the following information:
Publication's name, year and date of issue.

Without limiting the generality of the foregoing, no portion of the content may be translated from English or French into any other language without the prior written permission of Canada Mortgage and Housing Corporation.

The information, analyses and opinions contained in this publication are based on various sources believed to be reliable, but their accuracy cannot be guaranteed. The information, analyses and opinions shall not be taken as representations for which Canada Mortgage and Housing Corporation or any of its employees shall incur responsibility.

Housing market intelligence you can count on

FREE REPORTS AVAILABLE ON-LINE

- Canadian Housing Statistics
- Housing Information Monthly
- Housing Market Outlook, Canada
- Housing Market Outlook, Highlight Reports – Canada and Regional
- Housing Market Outlook, Major Centres
- Housing Market Tables: Selected South Central Ontario Centres
- Housing Now, Canada
- Housing Now, Major Centres
- Housing Now, Regional
- Monthly Housing Statistics
- Northern Housing Outlook Report
- Preliminary Housing Start Data
- Renovation and Home Purchase Report
- Rental Market Provincial Highlight Reports *Now semi-annual!*
- Rental Market Reports, Major Centres
- Rental Market Statistics *Now semi-annual!*
- Residential Construction Digest, Prairie Centres
- Seniors' Housing Reports

Get the market intelligence you need today!

**Click www.cmhc.ca/housingmarketinformation
to view, download or subscribe.**

CMHC's Market Analysis Centre e-reports provide a wealth of detailed local, provincial, regional and national market information.

- **Forecasts and Analysis –**
Future-oriented information about local, regional and national housing trends.
- **Statistics and Data –**
Information on current housing market activities — starts, rents, vacancy rates and much more.

Are you interested in housing research?

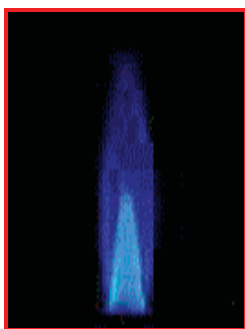
Stay up-to-date with the latest housing research findings and events related to sustainable housing and communities, housing conditions and trends, housing finance and more.

Subscribe Today to CMHC's Housing Research E-Newsletter!





LNG Information Paper No. 1



Basic Properties of LNG

GIIGNL's Technical Study Group has overseen the development of this Information Series of seven papers to provide factual information about Liquefied Natural Gas (LNG). In French, Spanish, Portuguese, or Italian speaking countries, the abbreviation GNL is used in place of LNG. This paper begins with a review of Basic LNG properties, which is a pre-requisite for accurately assessing potential LNG safety hazards and risks. For more information on these topics, additional references and weblinks are provided at the end of this paper.

Introduction:

A basic knowledge of LNG must begin with an examination of its chemical and physical properties. Chemical and physical properties are fundamental to understanding LNG correctly. The very properties which make LNG a good source of energy can also make it hazardous if not adequately contained. These properties determine how LNG behaves, affect our predictions about its behaviours, and influence how we assess and manage safety risks. Furthermore, to accurately understand and predict LNG behaviour, one must clearly distinguish its properties as a liquid from its properties as a gas or vapour.

The reader will note that discussions of the properties of LNG often contain ominous caveats like "depending upon its exact composition" because such specifics matter. It is inexact and inappropriate to make universal generalisations about LNG. It is especially important to be clear in thinking through how LNG would behave if accidentally or intentionally released (e.g., from a terrorist attack), because the outcome would be profoundly influenced by the actual situation and site-specific conditions.

Misunderstanding LNG is not uncommon and is often caused by confusion, incomplete, or inaccurate information about LNG properties. Since properties determine behaviour and influence how we manage

potential safety hazards and risks, having an accurate understanding is key.

A number of LNG companies have made commitments to educate the general public about their product. For example, companies in Japan and South Korea have gone to great lengths to share information about their facilities with the local communities and to educate them about LNG. For example, Osaka Gas Company and Tokyo Gas Company have installed Gas Science Museums at each of their terminals; the first one opened in 1982. More than 50,000 children, among other visitors, tour the museums every year and are able to observe table-top demonstrations of LNG properties and behaviours.

LNG is natural gas which has been converted to liquid form for ease of storage or transport. LNG takes up about 1/600th of the volume of natural gas. Depending upon its exact composition, natural gas becomes a liquid at approximately -162°C (-259°F) at atmospheric pressure.

LNG's extremely low temperature makes it a cryogenic liquid. Generally, substances which are -100°C (-48°F or less) are considered cryogenic and involve special technologies for handling. In comparison, the coldest recorded natural temperatures on earth are -89.4°C (-129°F) at the height of winter in Antarctica and the coldest reported temperature in a town was recorded in

Oymyakon (Sakha Republic) during Siberian winter (-71.2°C; -96.16 °F). To remain a liquid, LNG must be kept in containers which function like thermos bottles – they keep the cold in and the heat out. The cryogenic temperature of LNG means it will freeze any tissue (plant or animal) upon contact and can cause other materials to become brittle and lose their strength or functionality. This is why the selection of materials used to contain LNG is so important.

LNG is odourless, colourless, non-corrosive, non-flammable, and non-toxic. Natural gas in your home may have been liquefied at some point but was converted into its vapour form for your use. The reason the natural gas you use in your home has a smell is because an odourising substance is added to natural gas before it is sent into the distribution grid. This odour enables gas leaks to be detected more easily.

Key liquid and gas properties for LNG are:

- Chemical Composition,
- Boiling Point,
- Density and Specific Gravity,
- Flammability, and
- Ignition and Flame Temperatures.

These properties are listed on Material Safety Data Sheets (MSDS's). [\[Control Right click here to View MSDS.\]](#)

Chemical Composition:

Natural gas is a fossil fuel, meaning it has been created by organic material deposited and buried in the earth millions of years ago. Crude oil and natural gas constitute types of fossil fuel known as “hydrocarbons” because these fuels contain chemical combinations of hydrogen and carbon atoms. The chemical composition of natural gas is a function of the gas source and type of processing. It is a mixture of methane, ethane, propane and butane with small amounts of heavier hydrocarbons and some impurities, notably nitrogen and complex sulphur compounds and water, carbon dioxide and hydrogen sulphide which may exist in the feed gas but are removed before liquefaction. Methane is by far the major component, usually, though not always, over 85% by volume. **Table 1** displays the chemical compositions of the hydrocarbon compounds which make up natural gas, and the volume ranges in which they may be present in LNG. Pipeline natural gas may contain small amounts of water vapour.

LNG is often confused with liquefied petroleum gas (LPG), which in turn is often incorrectly identified as propane. In fact, LPG is a mixture of mainly propane and butane gases that exist in a liquid state at ambient temperatures

Table 1. Typical chemical composition of LNG (Source: Center for Energy Economics, www.beg.utexas.edu/energyecon/lng)

Chemical	Chemical Formula	Low	High
Methane	CH ₄	87%	99%
Ethane	C ₂ H ₆	<1%	10%
Propane	C ₂ H ₈	>1%	5%
Butane	C ₄ H ₁₀	>1%	>1%
Nitrogen	N ₂	0.1%	1%
Other Hydrocarbons	Various	Trace	Trace

when under moderate pressure (less than 1.5 MPa or 200 psi). In the U.S., Canada, and Japan, LPG consists primarily of propane (**Table 2**; next page). However, in many European countries, the propane content in LPG can be as low as 50% or less. Moreover, in some countries, LPG may contain a substantial portion of propylene.

LPG's differing composition and physical properties (from LNG) make its behaviour different as well. The propane and butane in LPG have different chemical compositions from methane, the primary hydrocarbon in natural gas and LNG. Propane and butane can be stored and transported as a mixture, or separately. Both are gases at normal room temperature and atmospheric pressure, like methane, readily vaporising. Propane liquefies much more easily than LNG (at -43°C [-46 °F] vs. -162°C [-259 °F] for LNG) so it is substantially easier to compress and carry in a portable tank. In fact, LPG is stored as a liquid under pressure at ambient (room) temperatures, whereas LNG is stored as a liquid only at very low temperatures and ambient pressure.

Boiling Point:

Boiling point is one of the most significant properties because it defines when gas becomes a liquid. Webster-Merriman on line (www.webster-merriman.com) defines “boiling point” as “the temperature at which a liquid boils” or at which it converts rapidly from a liquid to a vapour or gas at atmospheric pressure. The boiling point of pure water at atmospheric pressure is 100°C (212 °F). The boiling point of LNG varies with its basic composition, but typically is -162°C (-259 °F).

When cold LNG comes in contact with warmer air, water, or the environment, it begins to “boil” at that interface because the surrounding temperatures are warmer than

Table 2. Typical composition of LPG in % by volume
(Source: <http://www.environment.gov.au/settlements/transport/comparison/pub/2ch10.pdf>)

Country	Propane	Butane
Belgium	50	50
France	35	65
Ireland	100*	100*
Italy	25	75
Germany	90	10
UK	100*	100*
Denmark	50	50
Greece	20	80
Netherlands	50	50
Spain	30	70
Sweden	95	5

* **NOTE:** In Ireland and the U.K., LPG may be 100% of either basic gas.

the LNG’s boiling point, as shown in **Figure 1. Table 3** shows the boiling points of water and common gases.

The liquefaction process cools natural gas to change it to a liquid which reduces the volume occupied by the gas by approximately 600 times. LNG is converted back into natural gas for distribution to industrial and residential consumers. The LNG regasification process warms the LNG and converts it back into its gaseous form.

Density and Specific Gravity:

Density is a measurement of mass per unit of volume and is an absolute quantity. Because LNG is not a pure substance, the density of LNG varies slightly with its actual composition. The density of LNG falls between 430 kg/m³ and 470 kg/m³ (3.5 to 4 lb/US gal). LNG is less than half the density of water; therefore, as a liquid, LNG will float if spilled on water.

Specific gravity is a relative quantity. The specific gravity of a liquid is the ratio of density of that liquid to density of water (at 15.6°C/60°F). The specific gravity of a gas is the ratio of the density of that gas to the density of air (at 15.6°C). Any gas with a specific gravity of less than 1.0 is

lighter than air (buoyant). When specific gravity or relative density is significantly less than air, a gas will easily disperse in open or well-ventilated areas. On the other



Figure 1. LNG “boiling” at atmospheric pressure and temperature (Source: OSAKA Gas Co. Ltd.)

Table 3. Boiling points of water and some common gases (Source: Adapted from the Engineering Toolbox online, www.engineeringtoolbox.com/boiling-points-fluids-gases-d_155.html)

Fahrenheit (degrees F)	Celsius (degrees C)	Occurrence
212	100	Water Boils
31	-0.5	Butane Boils
-27	-33	Ammonia Boils
-44	-42	Propane Boils
-259	-162	LNG Boils
-298	-183	Oxygen Boils
-319	-195	Nitrogen Boils
-422	-252	Hydrogen Boils
-454	-270	Helium Boils
-460	-273	Absolute Zero

NOTE: Absolute zero is the coldest temperature theoretically possible, and cannot be reached, by artificial or natural means. By international agreement, absolute zero is defined as precisely 0 K on the Kelvin scale and is equivalent to $-273.15^{\circ}\text{C}/-459.67^{\circ}\text{F}$.

hand, any gas with a specific gravity of greater than 1.0 is heavier than air (negatively buoyant). The specific gravity of methane at ambient temperature is 0.554, therefore it is lighter than air and buoyant.

Under ambient conditions, LNG will become a vapour because there is no place on earth with a temperature of -162°C (-259°F). As LNG vaporises, the cold vapours will condense the moisture in the air, often causing the formation of a white vapour cloud until the gas warms, dilutes, and disperses as shown in **Figure 2**.



Figure 2. LNG vapour cloud created for training at Texas A&M in the LNG Live Fire Training Workshop (2005) (Photograph by A.H. Walker)

For a relative humidity of higher than 55%, the flammable cloud is totally included in the visible vapour cloud. If the relative humidity is less than 55%, the flammable cloud can be partially or completely outside of the visible cloud, which means that the vapours could be ignited even though the ignition source is distant from the visible vapour cloud. The size of the vapour cloud will depend on wind speed, direction, and other weather conditions and can easily be predicted by the appropriate related calculations. These very cold vapours will rise as they are sufficiently warmed by ambient air.

LNG vapours at the boiling point temperature ($-162^{\circ}\text{C}/-259^{\circ}\text{F}$) and atmospheric pressure have a relative density of about 1.8, which means that when initially released, the LNG vapours are heavier than air and will remain near the ground. However as methane vapours begin to rapidly

warm and reach temperatures around $-110^{\circ}\text{C}/-166^{\circ}\text{F}$, the relative density of the natural gas will become less than 1 and the vapours become buoyant.

At ambient temperatures, natural gas has a specific gravity of about 0.6, which means that natural gas vapours are much lighter than air and will rise quickly. Cold LNG vapours (below $-110^{\circ}\text{C}/-166^{\circ}\text{F}$) are negatively buoyant and more likely to accumulate in low areas until the vapours warm. Therefore, a release of LNG that occurs in an enclosed space or low spot will tend to replace the air (and oxygen) and make the area a hazard for breathing.

The rate of LNG vapour ascent depends upon the quantity of LNG released, ambient weather conditions, and where the LNG is released, e.g., confined or unconfined, low or elevated area, on land or on water. One strategy to manage the vapours is to create a downwind water curtain which helps block and/or divert the vapours away from possible ignition sources until the vapours warm and become buoyant, and/or dilute to a lesser concentration outside the flammable limits, which are discussed in the next section.

Heat input to LNG in any form will enhance vaporisation and dispersion. Such heat may be transferred from passive sources such as atmospheric humidity (which is a significant source), the ground or spill catchment areas, impoundments, pits and structures. LNG vaporises up to five times more quickly on water than on land, depending upon the soil condition. In fact, another strategy for managing the flammability hazard of LNG vapours is to use a water hose to warm the liquid more quickly (while avoiding contact with the super-cold LNG), increase vaporisation rates, and make the vapours buoyant sooner, rising away from ignition sources at ground level.

Flammability:

Flammability is the property which makes natural gas desirable as an energy source, and yet for the same reason flammability can be a safety hazard. It is very important to be clear: natural gas is flammable but LNG (the liquid form of natural gas) is not because of the lack of oxygen in the liquid. Since LNG begins vaporising immediately upon its release from a container, the important issue is when will the vapours be flammable and for how long?

Flammability Limits

Three things are needed to support a fire:

- A source of fuel (e.g., flammable gas or vapour),
- Air (oxygen), and
- A source of ignition (e.g., spark, open flame, or high-temperature surface).

This is known as the fire triangle (**Figure 3**). Several factors are required to start a fire from LNG vapours. In particular, the fuel and the oxygen have to be in a specific range of proportions to form a flammable mixture.

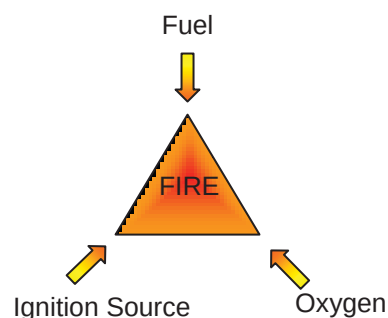


Figure 3. The fire triangle

This "Flammable Range" is the range of a concentration of a gas or vapour that will burn if an ignition source is introduced. The limits are commonly called the "Lower Flammable Limit" (LFL) and the "Upper Flammable Limit" (UFL) (**Figure 4**).

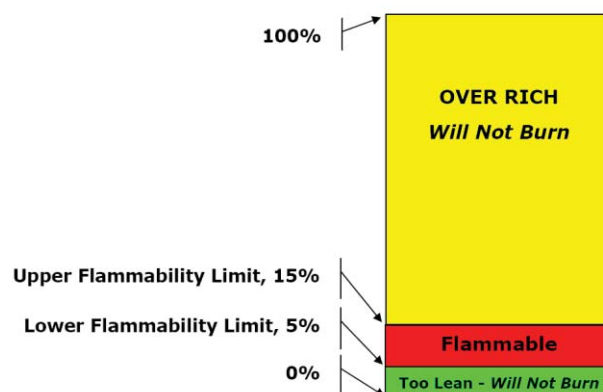


Figure 4. Flammability range for methane (Source: Foss 2003)

The LFL for methane is 5% and the UFL is 15% both by volume in air. Outside of this range, the methane/air mixture is not flammable. **Table 4** shows flammability limits for methane compared with other fuels. Many materials around us are flammable and it is important to be aware of each substance's flammability limits to assure safe handling and use. Materials that

Table 4. Flammability limits of hydrocarbon fuels (Source: NPFA Fire Protection Handbook)

Fuel	LFL	UFL
Methane	5.0	15.0
Butane	1.86	7.6
Kerosene	0.7	5.0
Propane	2.1	10.1
Hydrogen	4.0	75.0
Acetylene	2.5	>82.0

have wide flammable ranges make them dangerous to emergency responders because there is a longer time that they are within the flammable limits. For example, hydrogen and acetylene have a very wide range and Acetylene can burn whenever vapours are from just over 2% to over 80% in air.

In a closed storage tank or vessel, the percentage of methane is essentially 100% (mostly liquid and some vapours). Any small leak of LNG vapour from a tank in a well-ventilated area is likely to rapidly mix and quickly dissipate to lower than 5% methane in air. Because of the rapid mixing, only a small area near the leak would have the necessary concentration to allow the fuel to ignite. All LNG terminals use several types of equipment on and around the storage tanks and piping throughout the facility to detect any unlikely leakages and combustible gas mixtures. This safety equipment is described in [Information Paper No. 6](#).

Ignition and Flame Temperatures

The ignition temperature, also known as auto-ignition temperature, is the lowest temperature at which a gas or vapour in air (e.g., natural gas) will ignite spontaneously without a spark or flame being present. This temperature depends on factors such as air-fuel mixture and pressure. In an air-fuel mixture of about 10% methane in air, the auto ignition temperature is approximately 540°C (1,000°F). Temperatures higher than the auto ignition temperature will cause ignition after a shorter exposure time to the high temperature. [Table 5](#) shows the auto-ignition temperature of some common fuels, indicating that diesel oil and gasoline will auto-ignite at substantially lower temperatures than LNG.

The precise auto ignition temperature of natural gas varies with its composition. If the concentration of heavier hydrocarbons in LNG increases (e.g., the methane portion

Table 5. Auto-ignition temperature of some fuels at standard conditions (Source: BV 2009)

	Natural Gas	Diesel Oil	Gasoline
Auto-ignition temperature	599°C	260-371°C	226-471°C

of the natural gas begins to evaporate or be removed from the mix), the auto ignition temperature decreases. In addition to ignition from exposure to heat, the vapours from LNG can be ignited immediately from the energy in a spark, open flame, or static electricity when they are within the flammable limits.

LNG has a very hot flame temperature. Simply stated it burns quickly and is a better heat source than other fuels, e.g., gasoline. The methane in LNG has a flame temperature of 1,330°C (2,426 °F). In comparison, gasoline has a flame temperature of 1,027°C (1,880 °F), which means LNG burns hotter. Also, LNG burns quickly, at a rate of about 12.5 m²/minute, compared to gasoline's burn rate of 4 m²/minute. LNG produces more heat when burning because its heat of combustion is 50.2 MJ/kg (21,600 Btu/lb), compared to that of gasoline which has a heat of combustion of 43.4 MJ/kg (18,720 Btu/lb). The combustion of LNG produces mainly carbon dioxide and water vapour. The radiant heat of an LNG fire is a frequent safety concern of government regulators and officials, and the public.

Key Points and Conclusions

In closing, the reader should remember the key points of this first information paper:

1. First, and most importantly, one must understand that the very properties which make LNG a good source of energy can also make it hazardous if not adequately contained. While LNG is predominately methane (about 87%-99%), its composition also includes small amounts other hydrocarbons. The specific chemical composition natural gas is a function of the gas source and type of processing. The chemical composition of the natural gas and the properties of its hydrocarbon components determine how LNG behaves, affect our predictions about its behaviours, and influence how we assess and manage safety risks. Misunderstanding LNG is not uncommon and is often caused by confusion, incomplete or inaccurate information about LNG properties. One also must clearly distinguish its properties as a liquid from its properties as a gas or vapour.

2. LNG, the liquid form of natural gas, is a fossil fuel, like crude oil other hydrocarbon-based forms of energy and products.
3. The “boiling point” of LNG is -162°C ; -259°F , which is considered a cryogenic temperature. At this temperature (somewhat depending upon its actual composition), LNG evaporates to convert from a liquid to a vapour.
4. Conversely, LNG becomes a liquid at these cryogenic temperatures (-162°C ; -259°F) at atmospheric pressure. As a liquid, it takes up about $1/600^{\text{th}}$ the volume of natural gas. Consequently, it is generally transported and stored in a liquid state.
5. LNG is odourless, colourless, non-corrosive, non-flammable and non-toxic.
6. While natural gas is flammable, LNG is not. The flammability limits of methane are such that any small leak of LNG vapour from a tank in a well-ventilated area is likely to rapidly mix with air and quickly dissipate. Large leaks and spills are essentially precluded by a plethora of leak-detection systems and similar safeguards (which are discussed in later papers).

In summary, the basic properties and behaviours of LNG warrant that it be considered as a desirable option which can be managed safely when evaluating the mix of energy sources.

Next, in **Information Paper No. 2**, we will explore the “LNG Process Chain”. Subsequent papers in this series will include a discussion of the many ways in which LNG safety is assured, through Multiple Safety Layers, all firmly based on a foundation of solid Industry Standards, Regulatory Compliance and Codes. These “safety layers” include several key components of the industry's Risk Management framework. Included among them are Primary and Secondary Containment, Control Systems which promote Operational Integrity; Protocols, Operator Knowledge and Experience (which are reinforced by comprehensive and ongoing training). A protective umbrella of Safeguard Systems, Separation Distances, and Contingency Planning further enhances safe management of LNG. A graphic illustration of these “Multiple Safety Layers” is reflected in the figure at the end of this paper.

References and Additional Resources

California Energy Commission -

www.energy.ca.gov/lng/safety.html

Center for LNG - www.lngfacts.org

Foss, Michelle. 2003. LNG Safety and Security. Center for Energy Economics at the Bureau of Economic Geology, The University of Texas at Austin – http://www.beg.utexas.edu/energyecon/lng/documents/CEE_LNG_Safety_and_Security.pdf

The International Group of Liquefied Natural Gas Importers (GIIGNL) website - www.GIIGNL.org

Information from Bureau Veritas (BV) report for GIIGNL, 2009

National Fire Protection Association (NFPA) 2008 Flammable and Combustible Liquids Code Handbook.

National Fire Protection Association (NFPA) 2008 edition. Fire Protection Handbook.

Raj, Phani K. 2006. Where in a LNG vapour cloud is the flammable concentration relative to the visible cloud boundary? NFPA Journal, May/June 2006.

Raj, Phani K., and F.J. Katulak. 2007. What is the Truth on the LNG Safety? Why does the Public Perception on LNG Safety and What Can Industry Do? In: 15th International Conference & Exhibition on

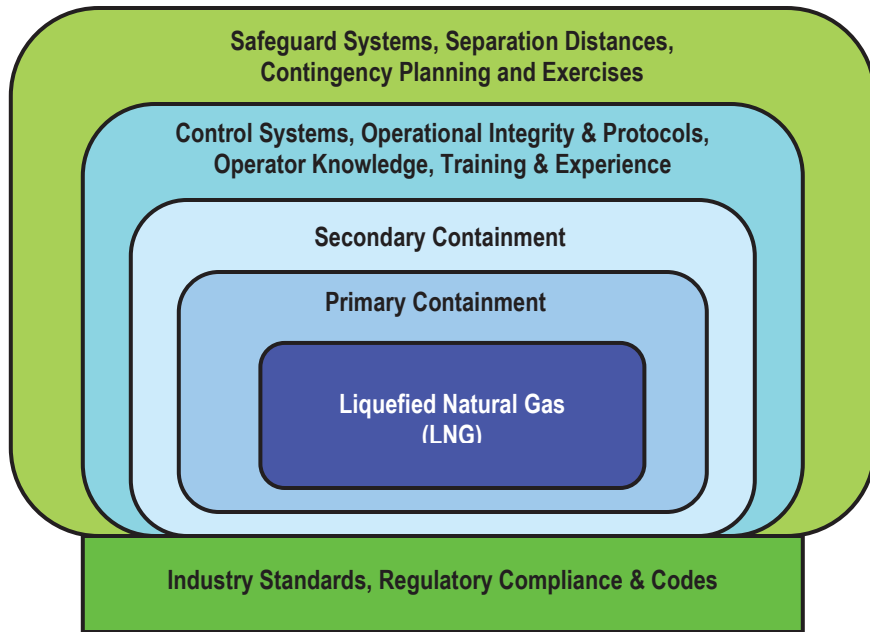
Liquefied Natural Gas, 24-27 April 2007. Barcelona, Spain.

Society of International Gas Tanker and Terminal Operators (SIGTTO) website - www.sigtto.org

Texas A&M. Texas Engineering Extension Service, LNG Live Fire Training Workshop and Resources Protection International. Student Manual. 2005 Brayton Fire Training Field, College Station Texas, USA. <http://www.altenergy.com/Technology/LPGProperties.htm>

US DOE - <http://fossil.energy.gov/programs/oilgas/storage/index.html>

Multiple Safety Layers Manage LNG Risk



The GIIGNL Technical Study Group has developed this 7-paper series to provide public readers with factual information about the LNG industry's multiple layers of safety, as illustrated in the figure to the left.

The GIIGNL Information Papers include:

- No. 1 – Basic Properties of LNG
- No. 2 – The LNG Process Chain
- No. 3 – LNG Ships
- No. 4 – Managing LNG Risks – Operational Integrity, Regulations, Codes, and Industry Organisations
- No. 5 – Managing LNG Risks – Containment
- No. 6 – Managing LNG Risks – Industry Safeguard Systems
- No. 7 – Questions and Answers (Q&A's)



For more information about these and other topics, or to obtain copies of this report series contact:

GIIGNL

1 rue d'Astorg
75392 Paris Cedex 08 (France)

Telephone: +33 01 49 24 74 74

Email: central-office@giignl.org

Web address: <http://www.giignl.org>

LNG

Liquefied Natural Gas A Strategy for B.C.'s Newest Industry



Message from the Premier

THE BC JOBS PLAN RELEASED IN SEPTEMBER is all about leveraging our competitive advantages to benefit British Columbians. Opening new markets for our exports, strengthening infrastructure to get our goods to market, and working directly with employers and communities will all help grow and strengthen our economy – creating jobs in every region of the province.

Building on our strengths is critical. So is breaking new ground. We've always relied on natural resources to fuel our economy. Now, with liquefied natural gas (LNG), we have a rare and exciting opportunity to build a whole new industry and use its development to spur other positive changes, such as growth in our clean-energy sector.

There will be challenges along the way. That is inevitable. It goes hand-in-hand with creating something new. As a government, we are committed to working closely with communities, First Nations and other important stakeholders. We are confident that, working together, we can reach our goals – investment, job creation and new economic opportunities – while protecting the environment and building a better quality of life for future generations.

With this LNG strategy, we are taking the next steps forward to harness British Columbia's strengths for the benefit of all our citizens. It's part of our plan to increase economic prosperity, create an environment where business and investment can flourish, and show the world that Canada really does start here.



Honourable Christy Clark
Premier of British Columbia

Global trade in LNG doubled between 2000 and 2010. It's expected to increase by another 50 per cent by 2020.

Message from the Minister



*Honourable Rich Coleman
Minister of Energy and Mines
and Minister Responsible for Housing*

1,000 cubic feet of natural gas costs under \$4 in North America in late 2011 – versus \$16 in Asia.

OVER THE NEXT 20 YEARS, GLOBAL DEMAND FOR NATURAL GAS

is expected to rise dramatically, fuelled by rapid economic growth in Asia. With the development of LNG – a shippable form of natural gas – B.C. is ideally positioned to compete for a share of that lucrative market.

Building a B.C. LNG industry will take time. And other jurisdictions – including the U.S., Australia and Africa – are also moving to develop their LNG potential. The good news is that B.C. is ready: we've been preparing for this opportunity for nearly a decade with progressive royalty programs, infrastructure upgrades, clean energy policies, comprehensive environmental assessments, and direct engagement with industry, First Nations and communities.

We are working hard to build our overseas markets through measures such as the Premier's recent trade mission to Asia. We are working with the industry to attract new capital and foreign investment. The federal government recently approved a 20-year export licence for the LNG facility being built in Kitimat – the first such licence ever issued in Canada.

With *The BC Jobs Plan*, the Province has committed to having our first LNG plant up and running by 2015, with a total of three LNG facilities operating by 2020. These are bold targets, but I am confident British Columbia will meet them.

Developing our LNG export potential is an excellent investment in our future. It will generate thousands of jobs and billions of dollars in new investment. That will mean more revenues for government to pay for services like health care and education. Equally important, it promises long-term stability for families and communities, with well-paying jobs, diversified economies and new opportunities to build expertise in a new global industry.

LNG Development – Our Vision for the Future

Quick Facts About Liquefied Natural Gas

- * LNG is natural gas, cooled to -160 degrees Celsius to keep it in a liquid form.
- * It is non-toxic, odourless, non-corrosive and less dense than water.
- * Compared to conventional natural gas, LNG takes up 600 times less space.
- * Unlike conventional natural gas, it can be shipped overseas, dramatically increasing its potential markets.
- * LNG has been safely used and transported around the world for 50 years.
- * It is a stable, low risk fuel.
- * If it spills, LNG will warm, rise and dissipate into the atmosphere.

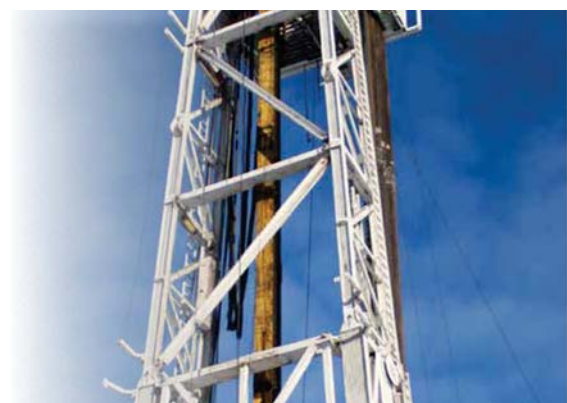
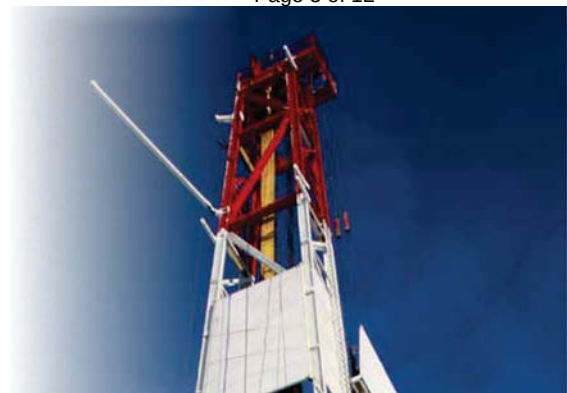
JUST A FEW YEARS AGO, PEOPLE WERE bracing for a shortage of natural gas in North America. Supplies of conventionally accessible gas were declining and contractors were considering options for importing liquefied natural gas – LNG – from other jurisdictions.

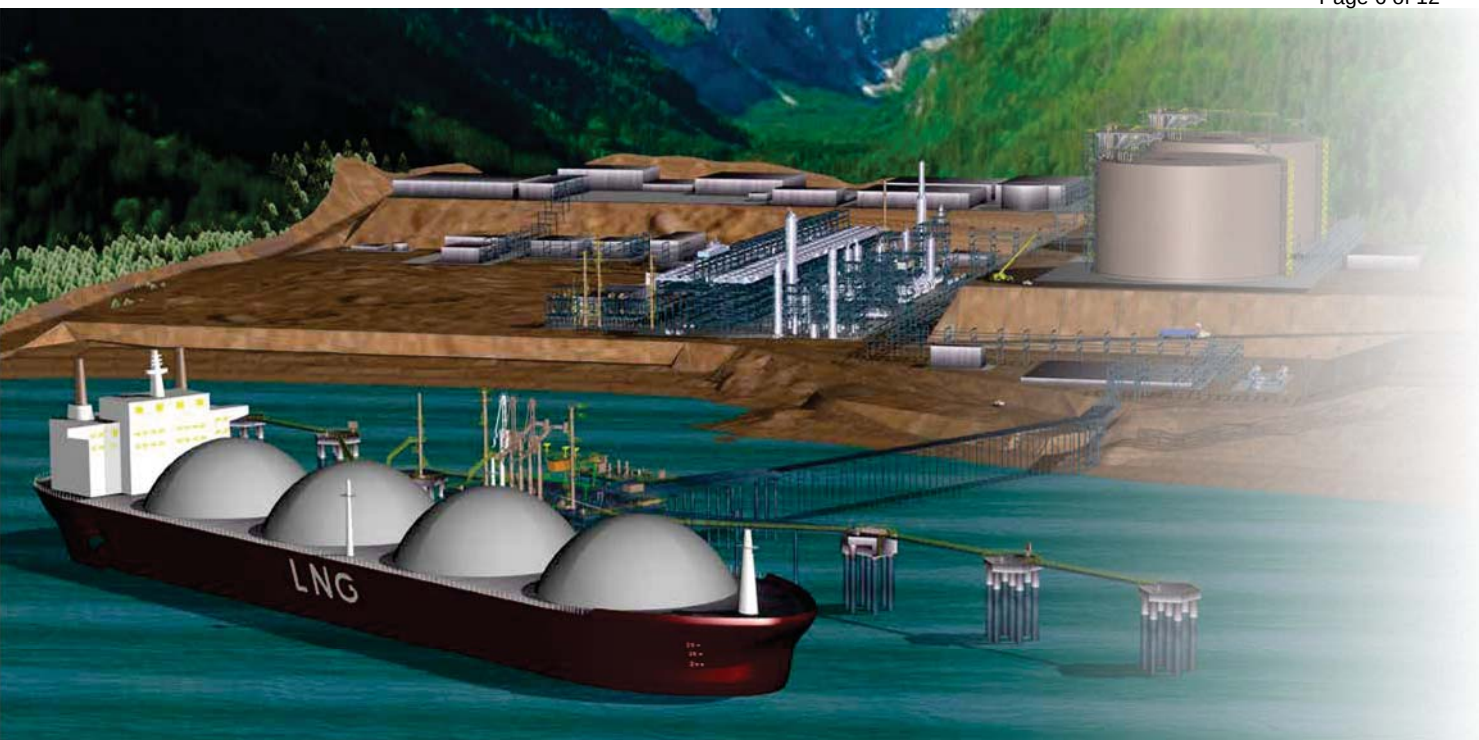
That all changed with the advent of technologies allowing for recovery of shale gas – an abundant form of natural gas with significant environmental benefits.

Natural gas is the world's cleanest-burning fossil fuel. For example, converting just one heavy duty truck from diesel to natural gas would reduce the same amount of greenhouse gas emissions as taking approximately 30 cars off the road. As proven supplies increase, so do the incentives to replace coal-fired generation with natural gas. So we believe it has an important role in the global transition to cleaner energy sources.

B.C. has been developing shale gas resources since 2005, generating billions of dollars in government revenue from land sales and royalties. Now we're moving forward to develop the potential of LNG for export.

Multiple investors across the natural gas sector have expressed interest in developing LNG export facilities. The first commercial LNG export facility in Canada is scheduled to open in Kitimat, on B.C.'s central coast, by 2015. And the Province has committed to working with interested investors, such as Shell Canada, to have three facilities in operation by 2020, assuming all environmental and permitting applications are granted.





Courtesy of Apache Canada LTD.

One of the first projects underway, the Kitimat LNG facility, has already earned federal and provincial environmental assessment approvals. It has strong support from the Haisla Nation, on whose land it's being built. And, in October 2011, it was granted the first-ever federal licence to export LNG from Canada. The Kitimat LNG plant will use clean electricity to liquefy natural gas, which results in lower emissions than plants elsewhere in the world.

Moving forward, additional LNG facility developments will use local clean energy with support from B.C.'s natural gas as necessary.

With this strategy, the Province intends to keep that momentum going, generating thousands of jobs and billions of dollars worth of new economic development to benefit families and communities in every part of British Columbia.

LNG: Generating Jobs and Revenues

The Province has committed to having three LNG facilities in operation by 2020, assuming all environmental approvals are granted. Based on current estimates from project proponents, that could mean:

- * over \$20 billion in direct new investment
- * as many as 9,000 new construction jobs
- * about 800 long-term jobs
- * thousands of potential spin-off jobs, and
- * over \$1 billion a year in additional revenues to government.

Vision: Three LNG plants in operation by 2020

Goals:

- * Keep B.C. competitive in the global LNG market
- * Maintain B.C.'s leadership on climate change and clean energy
- * Keep energy rates affordable for families, communities and industry

1. Keep B.C. competitive in the global LNG market



NATURAL GAS IS ONE OF B.C.'S MOST ABUNDANT RESOURCES, with vast untapped reserves throughout the northeast. Fears of a North American shortage disappeared in recent years with the advent of technologies making shale gas accessible. And while that has been a significant economic driver and revenue generator for our province, increased supply across North America has led to lower prices.

Natural gas will continue to be an important fuel for British Columbians, heating our homes, powering industry, and fueling our vehicles with fewer emissions than oil, gasoline or diesel. Developing liquefied natural gas for export will allow B.C. to dramatically expand its markets – and meet growing demand in Asia.

B.C. currently produces 1.2 trillion cubic feet (Tcf) of natural gas per year. Meeting our LNG development goals could add another 1.9 Tcf per year.



China and Japan are both pursuing new supply – China to fuel its massive modernization, and Japan to diversify its fuel supply. With demand growing quickly, prices in Asia are also up to four times higher than they are in North America.

All of this adds up to a great opportunity. But B.C. is not alone in pursuing it. Asian demand is fuelling a global race for long-term contracts to supply LNG, and B.C. faces stiff competition from jurisdictions such as Australia, the U.S., Qatar and Africa.

B.C.'s LNG Advantages

B.C. is well positioned to compete for a share of the lucrative Asian LNG market. Our advantages include:

- * lower shipping costs, thanks to our proximity to Asia
- * secure, stable government
- * vast natural gas reserves
- * high environmental standards
- * potential to access clean electricity
- * positive relationships with First Nations peoples
- * a well-established service sector
- * strong, updated regulations.

The Kitimat plant is on target to be fully operational by 2015 and several other projects are at the proposal stage. Recognizing that time is of the essence, the Province is taking an aggressive approach to developing the sector:

- an efficient regulatory system for LNG growth has been established
- overseas marketing is ramping up, supported by the New West Partnership with Alberta and Saskatchewan
- work is underway to streamline federal and provincial environmental assessments to create a single, more efficient process
- approaches to collaborative solutions for natural gas pipeline development are being explored, and
- collaboration with local communities, First Nations, industry and other levels of government is being strengthened to define more effective working relationships that benefit the entire province.

Next steps in helping to ensure B.C. has a competitive edge in this new global market will include investments in skills training. The Province is working with industry to define its needs and to help ensure the B.C. post-secondary system can deliver the targeted training needed to develop LNG, and to support the broader B.C. oil and gas sector.

2. Maintain B.C.'s leadership on climate change and clean energy

LNG – Helping to Address Global Climate Change

LNG development in B.C. will have lower life cycle green house emissions than anywhere else in the world by promoting the use of clean electricity to power LNG plants.

Natural gas has a key role to play in reducing greenhouse gas emissions (GHGs), and that is one of the driving factors behind its growing use in Asia. B.C. exports of LNG will significantly lower global GHG production by replacing coal-fired power plants and oil-based transportation fuels with a much cleaner alternative.

These reductions will affect B.C.'s own climate action targets, but since climate change is a global phenomenon, they will have a positive overall impact. Because other countries, including China, have their own GHG reduction goals, cleaner-burning LNG is even more attractive.

BRITISH COLUMBIA HAS A LONG HISTORY OF clean energy leadership, dating back to the 1960s when BC Hydro was established. Today, clean hydroelectric power, along with other renewable sources such as wind power and biomass, meets over 93 per cent of British Columbia's electricity needs. We are also offsetting two-thirds of our electricity demand growth through efficiency and conservations measures.

B.C.'s commitment to clean energy is also supported by the landmark Climate Action Plan, the first and most ambitious of its kind in North America.

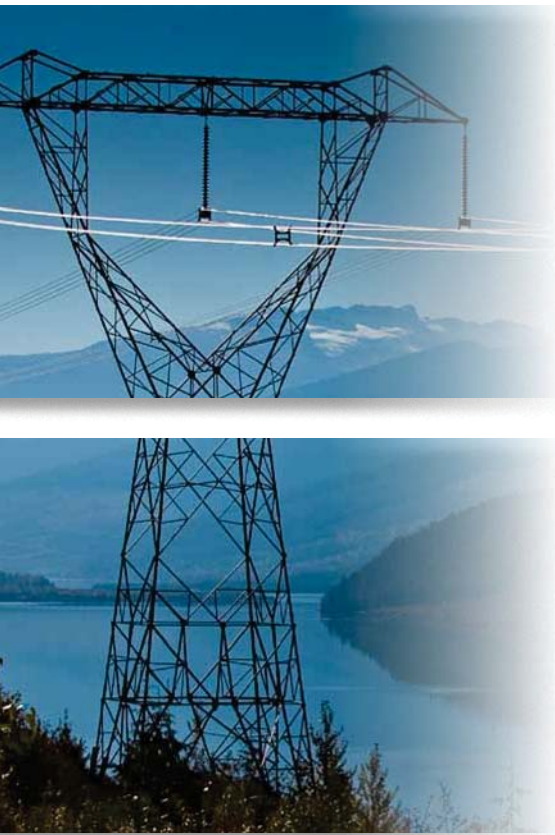
As part of the Jobs Plan, the Province is examining ways to grow the market for natural gas as a transportation fuel, in both CNG (compressed natural gas) and LNG forms. These alternatives can replace diesel in heavy duty fleets and other vehicles, and thereby help to lower GHG emissions.

At the same time, energy is needed to produce higher volumes of natural gas, and to operate LNG production plants. The first two LNG plants – BC Douglas Channel and Kitimat LNG – will use clean electricity to drive the liquefaction process, the first LNG plants to do so in the world. As a result, LNG development in British Columbia will have lower lifecycle greenhouse gas emissions than anywhere else. This will differentiate us in the global LNG export market.

As part of this strategy, the Province and BC Hydro will continue to work with the industry, First Nations, and with clean-energy producers to develop reliable, sustainable sources of supply.



Converting just one heavy duty truck from diesel to natural gas would reduce the same amount of GHG emissions as taking approximately 30 cars off our roads.



“With BC Hydro, our government is planning to meet the power demands required by new LNG facilities. LNG expansion will not be held back by a lack of supply of electricity.”

—Canada Starts Here: The BC Jobs Plan

3. Keep energy rates affordable

LIKE MOST MAJOR INDUSTRIES, LNG PRODUCTION REQUIRES a

steady source of power. In some cases, that could mean building new transmission lines or other types of infrastructure. That, in turn, has the potential to affect BC Hydro rates – and the Province is committed to ensuring the impacts on families and industry are minimized.

BC Hydro and the Province are currently working with LNG proponents to assess their future electricity needs – recognizing the key priority of keeping rates affordable. To offset the increased expense of operating new LNG facilities in the province, Government will ensure LNG developers contribute capital for infrastructure development and to the electricity supply required to serve each operation.

Another measure protecting consumers stems from a recent review of BC Hydro. That has led to changes in how government will implement its electricity self-sufficiency policy. This policy framework was originally implemented under the 2007 *Energy Plan* when economic growth was strong, natural gas prices were high and other jurisdictions were putting a price on carbon through taxes and planned cap and trade. Since that time, BC Hydro’s operating environment has changed, with market electricity prices dropping significantly as a result of the slow economic recovery, low natural gas prices, and the over building of subsidized renewable energy in the United States.

The original self-sufficiency policy required BC Hydro to acquire new electricity supply assuming that inflows into provincial water reservoirs would be at historically low levels, and to acquire an additional 3,000 gigawatt-hours of “insurance” by 2020. Moving forward, BC Hydro will plan electricity needs based on average water conditions, and the insurance requirement will be removed. Future demand from industrial development will now drive the need to purchase additional power.

These changes will enhance BC Hydro’s ability to optimize its unique and flexible hydro-based system and transmission connections to the western electricity market, creating more opportunities to earn income through short-term trading for the benefit of ratepayers.

The BC Hydro Review concluded that the impact of moving to average water and removing the insurance requirement would reduce electricity rate increases over the medium and long-term – up to eight per cent by 2016 and 20 per cent by 2020. This new policy direction will ensure that B.C. families and businesses will continue to enjoy some of the lowest electricity rates in North America, even as the government continues moving forward to implement the *Jobs Plan*.

Conclusion

LNG IS A BRAND NEW INDUSTRY WITH MASSIVE POTENTIAL for British Columbia. We have the supply, we have the technology, we have a great geographic advantage and, as we move forward to develop this industry, the whole province will benefit.

Thousands of people will have new jobs. Local economies will be more diversified. New skills training will be developed with new opportunities for future generations.

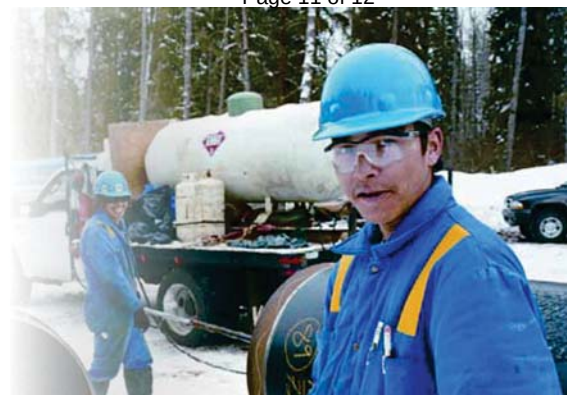
The LNG industry will generate economic spinoffs in areas such as the service sector and clean-energy development. First Nations will have new sources of economic strength and stability. And the Province will receive more revenues to pay for public services.

With this strategy, the government has laid out its critical priorities for LNG development:

- keeping B.C. competitive in the global LNG market;
- maintaining B.C.'s leadership on climate change and clean energy, and
- keeping energy rates affordable for families, communities and industry.

These three priorities will guide us going forward and help us to establish a thriving, competitive LNG industry that sets new standards for environmental and social responsibility.

As part of *The BC Jobs Plan*, this strategy is all about using our strengths to defend and create jobs in every community. This is B.C.'s time to lead and, together, we will.



Courtesy of TransCanada

“Not only have our people received immediate benefits from the project, in the form of a \$56 million payment for the sale of our equity in Kitimat LNG, but the long-term, regular lease and property tax payments combined with the employment and business opportunities associated with the project provide a greater measure of economic stability than we have ever experienced.”

— Former Haisla Nation Chief Counsellor Dolores Pollard
March 9, 2011



THE BC
JOBS PLAN



Natural Resources
CanadaRessources naturelles
Canada

Canada

Natural Resources Canada

Dollars to \$ense Energy Management Workshops

Since 1997, over 23,000 representatives of Industrial, Commercial and Institutional organizations from across Canada have enrolled in Dollars to \$ense workshops offered by Natural Resources Canada's Office of Energy Efficiency. There are six workshops to choose from:

- [Energy Management Information Systems](#)
- [Recommissioning](#)
- [Energy Management Planning,](#)
- [Spot the Energy Savings Opportunities,](#)
- [Energy Monitoring,](#)
- [Energy Efficiency Financing.](#)

The energy-saving tips offered at these workshops will help your organization:

- lower operating and production costs
- improve your competitive position
- reduce greenhouse gas emissions
- increase operational efficiency
- create a better work environment

Date Modified: 2011-10-25

Natural Resources
CanadaRessources naturelles
Canada

Natural Resources Canada

Benchmarking

The Canadian Industry Program for Energy Conservation (CIPEC) and its associated partners have developed two benchmarking programs for Canada's industrial sectors:

- [energy performance](#) benchmarking
- [best practices](#) benchmarking

Similar industrial companies may compare their energy use, greenhouse gas emissions and best practices using standard indicators for their industrial sector.

Energy Performance Benchmarking

Energy performance benchmarking focuses on a comparative analysis of energy use per unit of physical production, otherwise known as energy intensity.

Typically, the steps involved in energy performance benchmarking are:

1. Determine your plant's energy intensity by fuel type. Our [energy use calculator](#) will make the job easier.
2. Compare your energy intensity with that of your industrial sector by referring to the appropriate [benchmarking guide](#). You may also want to refer to the sector energy data available through the [Canadian Industrial Energy Efficiency Data and Analysis Centre \(CIEEDAC\)](#), at Simon Fraser University and supported by Natural Resources Canada.
3. Access the [technical information](#) available that will help you improve your facility's energy performance.
4. Select your CIPEC sector task force to obtain additional information specific to your sector.

Energy Performance Benchmarking



Energy Best Practices Benchmarking



Best practices benchmarking involves comparing operations and systems within your facility to best-in-class operations.

Generally, the steps involved in energy best practices benchmarking include:

1. Identify the areas for improvement that will benefit most from the benchmarking study.
2. Research and identify the key factors and variables you will use to measure the above.
3. Determine if the data is already available or how it will be obtained.
4. Analyze the data and identify the best practice/performance by selecting the best-in-class category (e.g., companies that perform each function at the lowest cost, with the highest energy efficiency).
5. Determine the conditions under which the best practices can be achieved and specify the action(s) that must be taken to achieve the desired results.
6. Implementation:
 - Set specific improvement targets and deadlines.
 - Develop a continuous process to monitor, review and update the data and analyze it over time. This will provide a basis for the monitoring, revision and recalibration of the measurements for future benchmarking studies.

Date Modified: 2011-07-08

-

✕

- [Yukon Energy's Blog](#)[RSS](#)[Facebook](#)[Twitter](#)

- **Current Energy Consumption**

-

-

-

-

-

-

-

-

-

-

- Hydro

-

- Diesel

See how current energy loads are being met by hydro vs diesel on our [Energy Consumption Charts](#).

- [Power Outage Updates](#)

[Home](#) > [Customer Centre](#) > [Tips & Tools](#) > Reading Your Bill

Reading Your Bill

Following is a sample Residential Customer bill. Roll over any of the circled numbers below for an explanation of what each item refers to.



[Riders are temporary adjustments on electricity bills that are in place until rates are revised through a General Rate Application.](#)



[The total you need to pay on this bill.](#)



[Pay your bill by this date to avoid a late charge.](#)

Project Number: K2033910

The California Evaluation Framework

**Prepared for the California Public Utilities
Commission and the Project Advisory Group**

June 2004

**TecMarket Works
And the project Team Members:**

**Megdal & Associates
Architectural Energy Corporation
RLW Analytics
Resource Insight
B & B Resources
Ken Keating and Associates
Ed Vine and Associates
American Council for an Energy Efficient Economy
Ralph Prah and Associates
Innovologie**



Project Number: K2033910

The California Evaluation Framework

Prepared for
Southern California Edison Company

To conduct a joint study supported by
**Pacific Gas & Electric Company
San Diego Gas & Electric Company
Southern California Edison Company
Southern California Gas Company**

As mandated by the
California Public Utilities Commission

June 2004

Submitted by
TecMarket Works Framework Team

**TecMarket Works
165 West Netherwood Road
Second Floor, Suite A
Oregon, WI 53575
Voice: (608) 835-8855
Fax: (608) 835-9490
E-mail: NPHall@TecMarket.net**

Acknowledgements

The authors wish to acknowledge and express our appreciation to the many individuals who contributed to the development of the California Evaluation Framework. Without the support and assistance of these individuals this effort would not have been possible.

The Project Advisory Group provided general and topic-specific project guidance and reviews of all chapters of the Framework. These individuals and their affiliations are:

- Marian Brown, Project Manager, Advisory Group Chairperson, Southern California Edison
- Eli Kollman and Jay Luboff, Energy Division, California Public Utilities Commission
- Don Schultz and Christine Tam, Office of Ratepayer Advocates, California Public Utilities Commission
- Mike Messenger and Sylvia Bender, California Energy Commission
- Valerie Richardson, Chris Ann Dickerson, and Kenneth James, Pacific Gas and Electric Company
- Athena Besa and Rob Rubin, Sempra Energy

In addition to the oversight and support of the Project Advisory Group, several individuals provided ongoing comments, support, and recommendations to the Framework as a whole and to individual chapters of the Framework. Over 90 individuals provided over 350 comments or recommendations to the Project Team. We wish to thank all those who provided comments. We have reviewed and discussed all of the comments received and have considered these comments in the development of the Framework. In many cases the comments were instrumental in guiding the development of the project's focus and the Framework chapters. Several comments required consultation with the Project Advisory Group to reach a consensus on the impact of the comment to the Framework. We wish to especially thank the following individuals who provided valuable comments that helped frame some of the discussions and improved the focus and structure of the Framework.

- Hayley Goodson, The Utility Reform Network
- Adrienne Kandel, California Energy Commission
- Patrick McCarthy and Bill Stiegelman, Aspen Systems
- Maureen McNamara, United States Environmental Protection Agency
- Robert Mowris, Robert Mowris and Associates
- Monica Nevius, Consortium for Energy Efficiency
- Dean Schiffman, Sempra Energy

- Phil Sissons, Sisson and Associates

We also want to thank the TecMarket Works Project Team that collaborated to design and develop the new California Evaluation Framework under the direction and oversight of the CPUC and the Project Advisory Group. These individuals include:

- Nick Hall, TecMarket Works
- Lori Megdal, Megdal & Associates
- Pete Jacobs and Stuart Waterbury, Architectural Energy Corporation
- Roger Wright, RLW Analytics
- Paul Chernick, Resource Insight Incorporated
- Ken Keating, Ken Keating and Associates
- Sharyn Barata, B&B Resources
- Ed Vine, Edward Vine and Associates
- Steve Nadel and Marty Kushler, American Council for an Energy Efficient Economy
- Ralph Prah, Ralph Prah and Associates
- John Reed, Innovologie

The Team and its authors would also like to thank Johna Roth of TecMarket Works for her editing and formatting work in preparing this document.

Table of Contents

ACKNOWLEDGEMENTS	I
CHAPTER 1: INTRODUCTION AND PURPOSE OF FRAMEWORK	1
PURPOSE OF EVALUATION AND A STANDARD EVALUATION FRAMEWORK	1
PROJECT APPROACH	2
EVALUATION TYPES AND CONSIDERATIONS COVERED BY THE FRAMEWORK.....	3
<i>Impact Evaluation, and Measurement and Verification Approaches (Chapters 6 and 7)</i>	<i>3</i>
<i>Process Evaluations (Chapter 8)</i>	<i>4</i>
<i>Information and Education Program Evaluation (Chapter 9).....</i>	<i>4</i>
<i>Market Transformation Program Evaluation (Chapter 10).....</i>	<i>5</i>
<i>Non-Energy Effects Evaluation (Chapter 11)</i>	<i>6</i>
<i>Uncertainty (Chapter 12).....</i>	<i>6</i>
<i>Sampling (Chapter 13).....</i>	<i>6</i>
<i>Evaluation and Cost-Effectiveness (Chapter 14).....</i>	<i>7</i>
EVALUATION RESEARCH NOT COVERED BY THE FRAMEWORK	7
THE USE OF EVALUATION ROADMAPS	9
CHAPTER 2: STUDY METHODOLOGY AND PUBLIC INPUT PROCESS	11
STUDY METHODOLOGY	11
<i>Development of Work Plan and Literature Review.....</i>	<i>12</i>
<i>Development of Principles and Roadmap.....</i>	<i>17</i>
<i>Development of Draft and Final Project Reports.....</i>	<i>17</i>
PUBLIC INPUT PROCESS	18
CHAPTER 3: USE OF THE EVALUATION FRAMEWORK BY DIFFERENT STAKEHOLDERS	21
CHAPTER 4: EVALUATION OVERVIEW AND ISSUES	27
PREFACE	27
CORE PURPOSE OF EVALUATION.....	27
THE TWO KEY FUNCTIONS OF EVALUATION – SUMMATIVE AND FORMATIVE EVALUATIONS.....	28
THE TWO BASIC TYPES OF EVALUATION	29
PROGRAM THEORY/LOGIC MODEL USE IN EVALUATION	30
<i>Developing a Program Theory</i>	<i>33</i>
<i>Reasons for an Evaluator to Use and Refine the Program Theory.....</i>	<i>35</i>
<i>Program Theories are Especially Important to the Evaluation of Complex Programs.....</i>	<i>36</i>
<i>The Relationship Between Program Theory, Market Assessment, and Market Theory.....</i>	<i>37</i>
EVALUATION ETHICS	38
HISTORY OF ENERGY EFFICIENCY PROGRAM EVALUATION IN CALIFORNIA	38
<i>1970s-1994: The Pre-Protocol Era</i>	<i>39</i>
<i>1994-1997: The Protocol Period</i>	<i>39</i>
<i>1998-2000: Energy Efficiency, Electric Industry Restructuring, and The CBEE Period</i>	<i>40</i>
<i>Post-2000: Transition Period and the Energy Crisis.....</i>	<i>41</i>
POLICY USE OF EVALUATION RESULTS	42
<i>Policy Based on Evaluation Results.....</i>	<i>42</i>
<i>Interpreting and Adjusting Results for Summative versus Formative Purposes</i>	<i>43</i>

<i>Making the Framework a Living Document</i>	<i>44</i>
APPENDIX TO CHAPTER 4.....	45
AN EXAMPLE OF A PROGRAM THEORY	45
<i>Program Activities</i>	<i>46</i>
<i>Program Assumptions.....</i>	<i>46</i>
AN EXAMPLE OF A LOGIC MODEL	48
<i>Complete Evaluation Ethics Document from the American Evaluation Association.....</i>	<i>49</i>
CHAPTER 5: UMBRELLA FRAMEWORK FOR EVALUATION PLANNING, FUNDING, AND RESEARCH PRIORITIZATION DECISIONS.....	53
INTRODUCTION	53
THE EVALUATION PLANNING CYCLE	54
COMPONENTS OF THE EVALUATION PLAN	56
REVIEWING AND APPROVING THE EVALUATION PLAN.....	57
EVALUATIONS TO FACILITATE AN INTEGRATED PLANNING PROCESS	58
<i>Input to Policy Considerations and Decisions.....</i>	<i>59</i>
<i>Early Feedback to Program Implementers</i>	<i>59</i>
<i>Program Lifecycle Stage.....</i>	<i>60</i>
<i>Evaluation Data Time Lags</i>	<i>60</i>
<i>Portfolio Planning Requirements.....</i>	<i>60</i>
<i>Evaluation Planning Requirements and Regulatory Oversight</i>	<i>61</i>
<i>Program Design/Solicitation, Selection/Review, and Implementation Preparation</i>	<i>61</i>
<i>The Value of Dispersed Timing for Evaluations</i>	<i>61</i>
<i>Contract Requirements for “Pay for Performance” Programs.....</i>	<i>62</i>
<i>Market Inertia</i>	<i>62</i>
<i>Timing Needs for Retention, Measure Life, and Technical Degradation Analyses.....</i>	<i>62</i>
<i>Oversight and Review</i>	<i>62</i>
EXAMPLE OF AN INTEGRATED EVALUATION TIMING SYSTEM	62
<i>Program Duration</i>	<i>65</i>
<i>Timing of Evaluation Activities.....</i>	<i>65</i>
<i>Multiple Offset Program Cycles</i>	<i>66</i>
<i>Impact Analysis for Final Contract Purposes.....</i>	<i>66</i>
<i>Program Planning Period.....</i>	<i>66</i>
ROADMAPS AND THE EVALUATION PLANNING PROCESS	67
<i>All Programs.....</i>	<i>68</i>
<i>Programs with Impact Goals.....</i>	<i>68</i>
<i>Information and Education Programs</i>	<i>70</i>
<i>Market Effects or Market Transformation Programs</i>	<i>71</i>
<i>Programs That Rely On Non-Energy Effects to Achieve Their Goals</i>	<i>72</i>
OVERVIEW OF THE FRAMEWORK	72
DETERMINING EVALUATION PRIORITIES.....	73
RECOMMENDATIONS FOR SETTING EVALUATION BUDGETS.....	75
EVALUATION SPENDING PRIORITIES	77
CONSOLIDATING PROGRAMS FOR EVALUATION.....	78
APPENDIX TO CHAPTER 5.....	81
AN EXAMPLE OF AN INTEGRATED PLANNING CYCLE.....	81
<i>Example 1 – Three-Year Program Cycle with Integrated Planning, Solicitation, and Evaluation Process.....</i>	<i>81</i>
<i>Example 2 – Offset Program Cycles</i>	<i>89</i>

CHAPTER 6: IMPACT EVALUATION	93
PREFACE	93
SKILLS REQUIRED FOR IMPACT EVALUATION	95
INTRODUCTION AND KEY ISSUES	96
<i>Overall Impact Evaluation Objective.....</i>	97
<i>Units of Measure.....</i>	99
<i>Interactions Between Impact Evaluation and Planning Processes.....</i>	99
<i>Selecting Methods.....</i>	100
THE BILLING ANALYSIS PATH	101
<i>Simple Aggregate Pre-Post Comparisons.....</i>	103
<i>Comparing Pre-Post Billing Data for Programs with Experimental Design</i>	104
<i>Types of Billing Analysis Models</i>	105
<i>Critical Billing Analysis Issues and Quality Control Efforts.....</i>	113
<i>Examples from Energy Efficiency Evaluation.....</i>	118
<i>Summary Guidelines</i>	120
ENGINEERING ANALYSIS	120
<i>Simple Engineering Models</i>	123
<i>Building Energy Simulation Models</i>	129
NET-TO-GROSS REQUIREMENTS AND METHOD OPTIONS.....	133
<i>Principles for Undertaking and Using Net-to-Gross Analysis.....</i>	134
<i>Selecting an Analysis Method</i>	135
<i>Survey-Based Methods.....</i>	136
<i>Econometric-Based Methods</i>	142
<i>Summary of NTGR Method Options and Roadmap Choices</i>	145
CHAPTER 7: MEASUREMENT AND VERIFICATION	147
PREFACE	147
<i>M&V and Process Evaluation.....</i>	147
<i>Skills Required for Measurement and Verification</i>	148
INTRODUCTION AND KEY ISSUES	148
<i>The Role of the International Performance Measurement and Verification Protocol</i>	148
<i>Statistical Context for M&V Studies</i>	150
UPPER LEVEL ROADMAP.....	150
<i>Measurement and Verification Options</i>	151
<i>Choosing the M&V Option</i>	152
M&V PLAN	154
<i>Identify Goals and Objectives</i>	155
<i>Specify Building Characteristics.....</i>	155
<i>Specify Data Products and Project Output.....</i>	157
<i>Specify M&V Option</i>	159
<i>Specify Data Analysis Procedures and Algorithms.....</i>	159
<i>Specify Field Monitoring Data Points</i>	159
<i>Estimate Data Product Accuracy.....</i>	159
<i>Specify Verification and Quality Assurance Procedures</i>	160
<i>Specify Recording and Data Exchange Formats</i>	160
MEASURE INSTALLATION VERIFICATION	160
OPTION A.....	160
<i>Data Resources.....</i>	164
<i>Uncertainty Analysis.....</i>	165
<i>Option A Example</i>	165
OPTION B.....	166
<i>Option B Example</i>	168
OPTION C	169

<i>Billing Data Collection</i>	171
<i>Comparison Models</i>	171
<i>Adjustment Factors</i>	171
<i>Analysis Techniques</i>	171
<i>Model Accuracy Criteria</i>	175
<i>Baseline Model Selection Criteria</i>	175
OPTION D.....	176
<i>Calibration Accuracy</i>	180
M&V ACCURACY REQUIREMENTS	181
METERING AND MONITORING	182
<i>Examples</i>	185
<i>Monitoring Duration and Schedule</i>	186
<i>Instrumentation</i>	188
<i>Measure Sampling</i>	193
<i>Sensor Placement</i>	194
<i>Measurement Uncertainty</i>	195
QUALITY CONTROL.....	195
DATA WAREHOUSE.....	195
<i>Data Resources</i>	197
<i>Data Review, Validation, and Analysis</i>	198
<i>Data Formatting</i>	198
<i>Database Update</i>	198
<i>Data Access</i>	198
APPENDIX TO CHAPTER 7: UNCERTAINTY CALCULATION EXAMPLE .	199
CHAPTER 8: PROCESS EVALUATION	205
PREFACE	205
SKILLS REQUIRED FOR PROCESS EVALUATIONS	206
INTRODUCTION AND KEY ISSUES	207
<i>Definition</i>	207
<i>Why These Two Goals are in the Definition</i>	207
<i>The Need for Process Evaluations</i>	210
<i>Process Evaluations are In-Depth Studies</i>	212
<i>Evaluation Timing</i>	212
<i>Process Evaluation Activities</i>	215
<i>Access to Program Materials, Resources, and Personnel</i>	219
MAKING DECISIONS AND SELECTING METHODS: THE PROCESS EVALUATION ROADMAP	220
<i>Steps in the Process Evaluation Roadmap</i>	221
PROCESS EVALUATION BUDGETS	228
CHAPTER 9: INFORMATION AND EDUCATION PROGRAM EVALUATION	229
PREFACE	229
INTRODUCTION AND KEY ISSUES	230
SKILLS REQUIRED FOR INFORMATION/EDUCATIONAL EVALUATIONS	233
IDENTIFYING THE ISSUES	233
MAKING DECISIONS AND SELECTING METHODS	234
<i>Program Evaluation Design, Tools, and Examples</i>	238

STEPS IN DECIDING IF AN INFORMATION/EDUCATION EFFECTS EVALUATION IS REQUIRED	242
CHAPTER 10: MARKET TRANSFORMATION PROGRAM EVALUATION.	245
PREFACE	245
SKILLS REQUIRED FOR MARKET EFFECTS EVALUATION	246
INTRODUCTION AND KEY ISSUES	249
MARKET IDENTIFICATION AND EVALUATION CONSOLIDATION	250
ASSESSING AND USING PROGRAM THEORY/LOGIC MODELS FOR MARKET EVALUATION	251
<i>Market Characterization and Baseline Measurement.....</i>	<i>254</i>
<i>Types of Market Indicators to be Studied</i>	<i>255</i>
<i>Orienting Theoretical Perspectives.....</i>	<i>257</i>
<i>Baseline Measurement: Some Practical Considerations</i>	<i>258</i>
MARKET INDICATORS AND MEASURING MARKET PROGRESS	259
<i>The Need to Tell a Story.....</i>	<i>260</i>
<i>The Need for Targeted Studies in the Early Years</i>	<i>260</i>
<i>The Usefulness of a Consistent Framework and Institutional Process.....</i>	<i>261</i>
<i>Dealing with Large Numbers of Market Indicators</i>	<i>262</i>
<i>Don't Forget About Gross Unit Savings.....</i>	<i>262</i>
ASSESS SUSTAINABILITY AND PLAN TO MEASURE LONG-TERM MARKET EFFECTS AND POST-PROGRAM SUSTAINABILITY	266
CHAPTER 11: NON-ENERGY EFFECTS EVALUATION	269
PREFACE	269
INTRODUCTION AND KEY ISSUES	269
<i>Definition</i>	<i>270</i>
<i>Why Non-Energy Effects are Included in the Framework</i>	<i>271</i>
<i>Evaluation Timing.....</i>	<i>272</i>
<i>Skills Needed to Conduct Non-Energy Effects Research</i>	<i>273</i>
<i>Evaluation Planning and Approach Tools.....</i>	<i>274</i>
RESEARCH METHODS AND APPROACHES	275
<i>Partial Listing of Non-Energy Effects Presented in the Literature.....</i>	<i>277</i>
MAKING DECISIONS AND SELECTING METHODS	279
<i>Steps in the Non-Energy Effects Evaluation Roadmap</i>	<i>279</i>
CHAPTER 12: UNCERTAINTY	287
INTRODUCTION AND KEY ISSUES	287
<i>Skills Required for the Uncertainty Analysis Tasks</i>	<i>287</i>
BIAS AND STATISTICAL PRECISION	288
<i>Basic Definitions.....</i>	<i>289</i>
<i>How Bias Arises.....</i>	<i>292</i>
<i>Special Techniques to Assess Bias</i>	<i>294</i>
<i>Reporting Potential Bias.....</i>	<i>297</i>
<i>Reducing the Standard Deviation of an Estimator.....</i>	<i>297</i>
INTEGRATING THE RESULTS FROM MULTIPLE EVALUATION STUDIES.....	298
<i>Estimating the Total Savings of a Portfolio of Programs</i>	<i>299</i>
<i>Contrasting the Results of Two Independent Studies</i>	<i>300</i>
<i>Pooling Two Statistically Independent Estimators of the Same Parameter</i>	<i>301</i>
<i>Chaining the Results of Two Evaluation Studies</i>	<i>303</i>
<i>Converting a P-Value to a Relative Precision or Error Bound</i>	<i>304</i>

ALLOCATION OF RESOURCES TO EVALUATION	305
<i>Bayesian Decision Theory and Power Analysis</i>	306
<i>Propagation of Uncertainty</i>	307
<i>The Fixed Relative Precision Criterion</i>	308
<i>Optimal Allocation for the Overall Savings of the Portfolio</i>	310
CHAPTER 13: SAMPLING	315
INTRODUCTION AND KEY ISSUES	315
<i>Skills Required for Sampling</i>	316
<i>Key Issues</i>	316
<i>General Steps in a Study</i>	317
<i>Population Database</i>	318
SIMPLE RANDOM SAMPLING	318
<i>Basic Definitions</i>	319
<i>Sampling Distribution of the Sample Mean</i>	320
<i>Expected Statistical Precision</i>	321
<i>Choice of Sample Size</i>	322
<i>Statistical Analysis</i>	324
<i>Applicability to Impact Evaluation</i>	325
STRATIFIED RATIO ESTIMATION	327
<i>Goals of the Section and Basic Definitions</i>	329
<i>The Ratio Model</i>	332
<i>Sampling Distributions</i>	334
<i>Expected Statistical Precision and Choice of Sample Size</i>	334
<i>Assessing the Error Ratio without a Prior Sample</i>	335
<i>Estimating the Error Ratio from a Sample</i>	336
<i>Model-Based Stratification</i>	336
<i>The Expected Statistical Precision for Any Sample Design</i>	337
<i>Applicability to Impact Evaluation</i>	338
A SAMPLING ROADMAP	340
<i>The Population Database</i>	340
<i>Sample Design</i>	342
<i>Sample Selection</i>	348
<i>Data Collection and M&V</i>	350
<i>Statistical Analysis</i>	352
<i>Reporting Guidelines</i>	359
APPENDIX TO CHAPTER 13.....	361
EXAMPLE AND USE OF STRATIFIED RATIO ESTIMATION FOR SAMPLING.....	361
<i>Choose the Sampling Approach</i>	364
<i>Sample Selection</i>	370
<i>Statistical Analysis</i>	375
<i>Analysis of Potential Selection Bias</i>	381
<i>Other Results</i>	383
CHAPTER 14: EVALUATION AND COST-EFFECTIVENESS.....	385
PREFACE	385
SKILLS NEEDED	386
COST-BENEFIT ANALYSIS FOR EVALUATORS.....	386
FACTORS INFLUENCING PROGRAM COST-EFFECTIVENESS.....	387
TREATMENT OF SPECIAL CATEGORIES OF COSTS.....	389
<i>Participant Supported Upgrades</i>	389
<i>Start-Up and Future Improvement Costs</i>	390

<i>Evaluation Costs</i>	390
UNDERSTANDING TEMPORAL AND GEOGRAPHICAL VARIATIONS IN AVOIDED COSTS.	391
PROVIDING USEFUL AVOIDED-COST ESTIMATION TO EVALUATORS	393
CHAPTER 15: OVERARCHING EVALUATION STUDIES	395
MEASURE SATURATION STUDIES	395
ENERGY-SAVINGS POTENTIAL STUDIES.....	396
PERSISTENCE AND RETENTION STUDIES	398
PORTFOLIO ANALYSES.....	400
BEST PRACTICE STUDIES	402
MARKET ANALYSES	402
<i>Market Structure and Operations Studies</i>	402
<i>Market Share Tracking Studies</i>	403
UPDATING OF KEY PARAMETERS	403
EVALUATION METHODOLOGY DEVELOPMENT	404
MANAGING AND FUNDING OVERARCHING STUDIES	404
APPENDIX A: GLOSSARY OF TERMS	407
APPENDIX B: BIBLIOGRAPHY	447
APPENDIX C: GUIDELINES FOR EVALUATION PLANNING	465
PROGRAM CLASSIFICATION	465
<i>Program Size</i>	468
<i>Program Strategies</i>	470
<i>Market Segments</i>	472
<i>Market Event</i>	473
<i>End Use or Measure Groups</i>	473
CPUC EVALUATION OBJECTIVES	475
<i>Measure Energy and Peak Savings</i>	475
<i>Measure Cost-Effectiveness</i>	475
<i>Provide Upfront Market Assessment and Baseline Analysis</i>	476
<i>Provide Ongoing Feedback and Guidance to the Program Administrator</i>	476
<i>Measure Indicators of Effectiveness and Testing Program Theory and Approach</i>	476
<i>Assess the Overall Levels of Performance and Success</i>	476
<i>Inform Decisions Regarding Compensation and Final Payments</i>	476
<i>Help Assess the Continuing Need for the Program</i>	476
EVALUATION PRIORITIES	478
<i>Risk to Portfolio Performance</i>	478
<i>Prior Evaluation History</i>	478
METHOD SELECTION FOR IMPACT EVALUATIONS	479
<i>Program Size</i>	479
<i>Market Event</i>	479
<i>Expected Impacts as a Fraction of Total Billing</i>	480
<i>Program Strategy</i>	480
<i>End Use/Measure Groups</i>	482
SAMPLING CRITERIA	488
DATA REQUIREMENTS	488

Chapter 1: Introduction and Purpose of Framework

The California Evaluation Framework (Framework) provides a consistent, systemized, cyclic approach for planning and conducting evaluations of California's energy efficiency and resource acquisition programs. This document presents that Framework and provides valuable information concerning when evaluations should be conducted, the types of evaluation that can be conducted, and a discussion of approaches for conducting those studies. The intended audience for various sections includes policy staff, program portfolio managers, program planners and implementers, evaluators, and other stakeholders. (Chapter 3 provides a review of the potential users of this document and guidance about which chapters and at what level of review they would be most beneficial.)

Purpose of Evaluation and a Standard Evaluation Framework

There are two primary purposes for conducting evaluations of energy programs in California. These are: 1) to reliably document program effects, and 2) to improve program designs and operations to be more cost-effective at obtaining energy resources. All program evaluation efforts associated with California's energy programs fall under one or both of these overall purposes for conducting evaluations.

To attain these goals and to allow policy makers to more completely understand the potential effects from programs and program portfolios, the California Public Utilities Commission (CPUC) mandated the construction of a statewide evaluation framework. This Framework provides a rigorous systems approach to conducting evaluations so that all programs are able to document their effects and be compared to other programs and supply options.

In order to initiate the construction of the Framework, the Project Advisory Group developed a request for proposals that was approved by the CPUC. Through a competitive bidding process, the group selected the TecMarket Works team to develop the Framework within a process allowing extensive public input. This document presents the results of that effort. The Framework provides important information on what types of program evaluations are sought and provides guidance on methods and approaches that can lead to high quality evaluation studies. The Framework is not a step-by-step evaluation procedures manual, but it is an evaluation guidance document that contains information relating to the focus and implementation of energy program evaluation studies in California.

There are several CPUC evaluation goals incorporated into this project that served to focus the efforts of the CPUC staff, the Project Advisory Group and the TecMarket Works team. These goals serve to provide an evaluation framework that when implemented in California will:

1. Provide reliable evaluation results to support energy policy and supply decisions,

2. Allow programs to be equably compared according to their energy impacts,
3. Help understand and verify program energy and peak savings,
4. Help identify and quantify market and non-energy effects,
5. Provide information needed to estimate program cost-effectiveness, and
6. Provide recommendations for program changes that help improve cost-effectiveness.

In addition to accomplishing the above high-level goals, the Framework was designed to accomplish or support the following more specific objectives:

1. Increase the level of reliability of program savings impact estimates for use in resource planning forums where the uncertainty of these estimates needs to be compared against the uncertainty of other key components of the resource plan.
2. Increase the quality of feedback to program administrators from evaluation projects to both improve program designs and increase the net savings from their programs.
3. Provide guidance to program administrators on what types of evaluation are recommended and are likely to be most beneficial for documenting operations and objective accomplishments.
4. Provide guidance to program administrators on the methodological approaches and study focus needed to perform specific types of evaluations.
5. Provide a Framework with flexibility that allows for the use of alternative evaluation approaches, especially when they can be shown to provide as reliable results as the methods presented in the Framework.

This document accomplishes these objectives and describes the recommended efforts needed to effectively evaluate different types of energy programs. The document also describes the skill sets necessary to accomplish these efforts. This document does not make recommendations on which organizations should be responsible for developing evaluation plans or completing the evaluation studies. The procedures for implementing this Framework are to be determined by the California Measurement Advisory Council (CALMAC) and the CPUC. Additionally, it is premature to settle the evaluation procedures until the overall framework for administering programs is determined by the CPUC.

Project Approach

In constructing this Framework, the TecMarket Works team relied to a substantial degree on past California evaluation protocols developed in the early 1990s and modified over the intervening years, and on other seminal evaluation publications found in the literature. These documents are presented in the bibliography section of this report. Readers are encouraged to reference these documents, as they are a critical part of the evolution of the

energy program evaluation field. The project team then began working with the CPUC staff, the Project Advisory Group and interested stakeholders through a project development process involving significant issue discussions with the Project Advisory Group and then incorporating extensive public input as the draft framework was developed. (See Chapter 2 for further information on the project methodology and public input process.)

Evaluation Types and Considerations Covered by the Framework

Nine key components of the Framework are incorporated into the Framework's design for performing program evaluations (See Chapter 5: Umbrella Roadmap). These nine components consist of the types of evaluations covered by the Framework and key concepts and considerations that go into planning and conducting evaluations in California. These nine components are summarized below and detailed in the chapters specifically dealing with each component.

Impact Evaluation, and Measurement and Verification Approaches (Chapters 6 and 7)¹

Impact evaluations focus on estimating the gross and net effects from the implementation of one or more energy efficiency programs. Most program impact projections contain ex-ante estimates of savings. These estimates are what the program is expected to save as a result of its implementation efforts. These estimates are used for program planning and contracting purposes and for prioritizing program funding choices.

The impact evaluation focuses on identifying and estimating the amount of energy and demand the program actually provides. Estimates of actual savings are ex-post savings; program savings that can be documented after the program has made the changes that are to produce the savings. Savings induced by the program are called "net" savings, as they are beyond or in addition to what would have occurred without the program. Ex-post net savings are the savings estimates as measured/verified as being achieved by the program.

In approving programs, the CPUC bases its decisions on ex-ante savings, the savings that are expected to be delivered by the program. What the state receives from the program is ex-post net savings, the estimated energy savings that are actually caused by the expenditure of the program dollars. These savings may change over time. Ex-post net savings, documented via an impact evaluation, can be in a range of effects that are from substantially more than the projected ex-ante savings to substantially less, or somewhere

¹ Impact evaluations and their supporting measurement and verification (M&V) activities are discussed in two chapters of the Framework. The decision to place these related efforts within two chapters (rather than one chapter with a M&V sub-chapter) is because impact evaluations can be (in some limited cases) conducted without supportive M&V efforts. In addition, the authors agreed that these topics are substantially different and are important enough that each of these topics needs to be treated as a separate chapter allowing readers to focus on one topic. They are discussed together here to indicate their importance in the impact estimation process.

in between. Over the years, impact evaluations have helped to refine and define ex-ante savings estimates so that today the difference between ex-ante savings and ex-post savings is expected to be small, unless there is a difference in the markets being served by a program, the vendors offering the program, the technologies promoted by a program, or other changes that can be expected to effect achieved savings. The impact evaluation documents in a rigorous estimation process the energy and demand impacts that the program provides. The Impact Evaluation chapter discusses the evaluation approaches suggested for program evaluators to document the amount of energy saved via the program's efforts. However, other methods and approaches can be used if they provide the same level of expected accuracy as those methods presented in the Framework.

The impact evaluation will often employ metering, monitoring, and verification tools to help accurately estimate the ex-post program savings. These efforts are typically referred to as "M&V," meaning either *Measurement and Verification* or *Monitoring and Verification*, depending on the publications or reference used. The M&V approaches typically used in impact evaluation are discussed in the M&V chapter of this document (Chapter 7), but in summary, are some form of field measurements taken to help identify how much energy is used before the program actions are taken, how much energy is being used after the actions are taken, the use conditions associated with an installed technology, or a change in behaviors that is to produce the energy savings.

Process Evaluations (Chapter 8)

The process evaluation is a systematic assessment of an energy efficiency program for the purposes of documenting program operations at the time of the examination and identifying improvements that can be made to increase the program's efficiency or effectiveness for acquiring energy resources. In addition, a process evaluation can also help increase the effectiveness of other programs by providing other program planners and administrators with the evaluation results. These planners can then review the process evaluation results to determine if their programs can benefit from the evaluation's findings and recommendations.

The Process Evaluation chapter discusses the intent and focus of the process evaluation and the skill levels needed to conduct these evaluations. It also identifies the variety of investigative issues associated with the process evaluation and the tools typically used in these studies. Additionally, the chapter discusses the need for process evaluations to be conducted in time for programs to benefit from the evaluation findings, and the need to establish early information feedback systems with the program administrators to allow for evaluation results to feed the program redesign process to obtain the maximum level of energy resources within the program delivery period. One of the best methods for identifying "best practices" is through the process evaluation.

Information and Education Program Evaluation (Chapter 9)

Information and education program evaluations focus on assessing the degree to which program goals are accomplished and estimating the effects of the program's activities on

their target markets. They can also serve as an information source for assessing the cost-effectiveness of the program. Evaluations of information and education programs in California typically have a different research goal than the evaluations of programs that have energy impact goals. Most of California's information and education programs do not have energy impact goals and are not expected to be cost-effective from an energy acquisition perspective. Instead these programs are designed to influence the ability of other programs to achieve their energy impact goals or they are focused on trying to influence the short-term or long-term decision processes associated with acquisition and use of energy-consuming technologies or behaviors. The evaluation of information and education programs within the Framework focuses on documenting the effects of the program at reaching their information transfer or educational goals.

Market Transformation Program Evaluation (Chapter 10)

The Market Transformation (MT) Program Evaluation chapter of the Framework focuses on the evaluation of program-induced market effects when the program being evaluated has a goal of making longer-term lasting changes in the way a market operates. These evaluations examine changes within a market that are caused, at least in part, by the energy efficiency programs attempting to change that market. These evaluations are challenging, as markets are constantly in a state of change as new and competing technologies are offered or as other non-program market transformation efforts compete with the program's efforts.

Two other forms of market evaluations (market baseline studies and market operations studies) are also briefly discussed in this chapter, as their development and use may be required to support a MT program evaluation.

Market baseline studies focus on documenting the status of the market that the program is attempting to change, so that changes caused by the program can be compared to the baseline conditions. The Framework discusses the challenges associated with these studies and the need to identify full or partial causal effects to the observed change in order to know which changes should be considered as normal baseline changes.

A market operations evaluation is a study that documents how a market operates relative to technology flows and information exchange within the targeted market. A primary purpose of these studies is to guide program theories and designs with factual information about how a market operates. This allows the program design to be integrated into the market in a way that supports accomplishing program goals. These studies also help in developing recommendations for program changes to make the program more successful at reaching its goals. From this perspective, it is easy to see why a market operations evaluation may be part of or integrated into a program process evaluation. Like the process evaluation, the market operations evaluation has the ability to help all programs that have impact goals within the market being assessed.

Non-Energy Effects Evaluation (Chapter 11)

Non-energy effects evaluations look at the intended or unintended effects that occur in addition to the energy impacts associated with a program. The Public Goods Charge (PGC) funds in California are obtained from ratepayers to support energy efficiency investments in lieu of supply-side investments for energy. Consequently, it is CPUC policy (as expressed by CPUC staff during this project) that calculations of program cost-effectiveness should be based only on the value of energy savings created by the program, and not on other benefits. Given this approach, non-energy effects evaluations conducted on programs funded through the Public Goods Charge are limited in scope. In general, non-energy effects evaluations can be conducted when the evaluation focuses on providing information that will help increase the energy impacts of a program or programs, or when the CPUC explicitly approves an evaluation in order to document specific non-energy effects that are of interest to the CPUC. This chapter identifies the conditions for conducting a non-energy effects evaluation and provides a discussion on the variety of types of non-energy effects that have been studied, and the methodologies that have been used.

Uncertainty (Chapter 12)

As a result of CPUC policies (expressed during this project) associated with this Framework, it is the explicit intent of this Framework that evaluations conducted on California energy efficiency programs be conducted in a way that provides reliable technology-specific, or in some cases, program-specific ex-post net energy impact findings.

The chapter describes how evaluations should assess and report the level of uncertainty and potential sources of bias associated with the evaluation findings and explain what actions are taken to mitigate these. Evaluation users should be able to determine the reliability of the study results and to determine if the results can be used for supply decisions, for public program policy making, or for updating deemed energy factors. The information is also designed for use in a summative fashion, where the propagation of uncertainty could be calculated for a group of programs, for the PGC-funded portfolio, or the overall portfolio of energy efficiency programs.

Sampling (Chapter 13)

This chapter presents and discusses the relationship between sample size and sample selection methods and the ability to assess ex-post effects that fairly represent the impacts that a program has achieved. Methods associated with calculating sample sizes and for allocating samples for studies are outlined in this chapter. The chapter also provides information for the calculation of relative precision, factors needed to create efficient sample designs, and how to use these within an evaluation's analysis phase.

Evaluation and Cost-Effectiveness (Chapter 14)

This chapter is targeted to three audiences in the area where issues overlap between program evaluations, cost-effectiveness analysis, and their uses and interpretations. First, it helps evaluators see how the results from their evaluations will be used in cost-effectiveness analysis. Second, it demonstrates to program staff and administrators who calculate or use cost-effectiveness analysis how evaluation and cost-effectiveness work together. Third, it helps policy makers understand some of the key issues involved in using evaluation results to estimate cost-effectiveness, since these tests are often used to inform a policy decision about whether to continue to invest in a program.

The chapter does not establish methods for calculating avoided costs or for conducting cost-effectiveness tests. The *California Standard Practice Manual* and the *Avoided Cost and Cost Effectiveness Study* address these topics.^{2, 3}

Evaluation Research Not Covered by the Framework

While the Framework covers a wide range of types of evaluation research, it is not designed to cover three types of tangential research that may be associated with designing, planning, or conducting evaluations of energy programs. The Framework does not address the following types of research efforts:

- Program-specific market research or market operations evaluations for purposes other than conducting independent evaluation assessments;
- Research, development, and deployment (RD&D) program evaluations;
- Low-income program evaluations; and
- Overarching evaluation studies.

A brief explanation of these exclusions is presented below.

Program-specific market research or market operations evaluations for purposes other than conducting independent evaluation assessments

It is often desirable for program administrators or others to conduct market research for the purposes of planning, designing, administering, or implementing one or more energy programs. When these studies are conducted for reasons other than to evaluate the performance of an energy program, the recommended approaches presented in the Framework do not apply. This is not meant to imply that this research is not important. It can be critical for the design of cost-effective programs or to optimize program design. As such, allocating program resources dollars for this type of research can be well-justified.

² *Standard Practice Manual (SPM) Economic Analysis of Demand-Side Management Programs.* (* California State Governor's Office 2001)

³ *California Avoided Cost and Cost Effectiveness Study* currently being conducted.

Research, development, and deployment (RD&D) program evaluations

This Framework does not cover evaluations of energy technology research, development, and deployment programs. These programs typically focus on the development and commercialization of new energy products or services. They are not specifically designed to acquire near-term energy savings. Because of this, this Framework does not cover them. Excluding evaluation of RD&D efforts from this Framework is not meant to convey that RD&D programs or evaluating RD&D programs is not important to the resource acquisition process in California. They are an important part of the long-term chain of events leading to the ability of energy programs to acquire energy resources. California has a long history of providing energy efficiency program funding for emerging technology programs. These programs are designed to speed market acceptance of emerging technologies that have successfully reached the commercialization phase but are not widely used due to customers' lack of knowledge, concerns, or questions about them. As a result, these programs can be evaluated using approaches similar to those for conducting evaluations of information and education programs.

Low-income program evaluations

California's low-income program evaluations are also not covered by this Framework. In California these programs have their own evaluation framework and evaluation decision policies specific to these equity programs. The offering of low-income programs is often influenced more by policies based on the need for the program's services than on the amount of energy they save or the cost-effectiveness of the programs.

Overarching evaluation studies

There are a number of other types of evaluation studies that are not program-specific in their focus or in the researchable issues they address that are not covered by this Framework. These studies are discussed in Chapter 15 as there are significant contributors to the overall evaluation effort for efficiency programs and can benefit multiple programs or program portfolios. The Framework does not attempt to guide the evaluation approaches of these types of evaluations. Evaluations that fall under the overarching classification include:

- Measure saturation studies,
- Energy-savings potential studies (technical, economic, achievable-market),
- Portfolio analyses (including "best practices" and "lessons learned" studies),
- Market and market operations analysis (beyond program level),
- Studies that update key parameters that influence multiple programs (e.g., measure lifetime, avoided costs), and
- Development of improved methodologies for evaluating programs.

However, when program-specific evaluations are consolidated, so that the evaluation deals with a group of programs, the Framework can guide and assist these evaluations.

The Use of Evaluation Roadmaps

Within each of the chapters dealing with a specific type of evaluation is an evaluation decision “roadmap.” These roadmaps are designed to assist the program administrator with their evaluation planning process and the related program-specific evaluation decisions. The administrators of California’s energy efficiency and resource acquisition programs are encouraged to use these roadmaps in their evaluation planning processes. The roadmaps consist of a set of decision trees or decision flow diagrams that walk through the process of determining if an evaluation is needed and what type of evaluations, methods, or decisions are expected.

Chapter 2: Study Methodology and Public Input Process

Study Methodology

The California Evaluation Framework (Framework) was developed to help program administrators plan, prioritize, and conduct their program evaluation, measurement, and verification (EM&V) activities. A key project goal incorporated into the development of the Framework was to produce a single “user-friendly” document that could provide evaluation guidance for a comprehensive set of program-specific evaluation efforts. This effort was to incorporate a wide range of public input into that process. The final Framework document is designed to serve as an evaluation roadmap, providing practical guidance that program administrators can follow in order to plan their evaluation efforts to meet the needs of the CPUC and other program or energy supply stakeholders.

The previous California framework study, *A Framework for Planning and Assessing Publicly Funded Energy Efficiency*, (hereafter referred to as the *2001 Framework Study*⁴) can be viewed as an authoritative text on market transformation programs and the evaluation of those programs, and should be considered a key supportive document to this Framework. The *2001 Framework Study* describes how market transformation programs can be evaluated and what tools can be used in those evaluations. It also provides a theoretical foundation for placing market interventions (whether from a market transformation or resource acquisition perspective) into energy efficiency markets.

However, this project is built upon an Advisory Group and California Public Utility Commission (CPUC) vision that was established early in the project, that program administrators need an evaluation road map that serves as a decision-guidance system for determining what research should be conducted for specific types of energy programs. This Framework also provides advice on how these evaluations should be conducted. The Framework is also designed as a reference manual for evaluators, providing guidance on planning, budgeting, and implementing program-specific evaluations. Additionally, the Framework can be used as an evaluation policy support document, providing policy makers with an information source from which evaluation policies and budgets can be guided. The Framework provides information to help guide the selection of evaluation approaches, including the measurement and verification approaches that must inform the impact estimation process. Likewise, the Framework reflects the needs of program administrators, designers, and policy makers to have early feedback from the evaluation process to support the evolution of program services to be increasingly cost-effective. At the same time, the Framework recognizes that it is equally important that the EM&V efforts should themselves be cost-effective.

⁴ *A Framework for Planning and Assessing Publicly Funded Energy Efficiency* (* Sebold et al. 2001). Primary investigators included Drs.: Frederick D. Sebold and Alan Fields, RER; Lisa Skumatz, Skumatz Economic Research Associates, Inc.; Shel Feldman, Shel Feldman Management Consulting, Miriam Goldberg, Xenergy, Inc.; Ken Keating, Consultant; and Jane Peters, Research Into Action, Inc.

Development of Work Plan and Literature Review

Project Initiation, CALMAC Kick-off, and Work Plan Development

The first steps in the development of the new Framework consisted of a project initiation meeting and the development of a detailed Draft Work Plan. These efforts began with a project initiation teleconference with the Project Advisory Group. This first meeting began to set the stage for the development of the Framework and to identify the evaluation subjects to be addressed in the Framework. Following the initiation meeting the project team conducted a series of meetings in which the draft work plans were developed, reviewed, and modified into the final project work plan.

To begin the Framework's public input process, a project kick-off public meeting was held in San Francisco shortly after the draft program plans were developed. An announcement for this meeting was publicly posted on the California Measurement Advisory Council (CALMAC) service list. The basic design of the project, the scope of the work, and the anticipated project efforts were presented at the public meeting and on the CALMAC web site. Public input on these presentations was received during the CALMAC meeting and via telephone and e-mail correspondence. These comments focused on the project's scope of work, issues to be addressed, the needs and potential uses of the project, and questions concerning the project's processes and the work to be performed. Over 150 comments were received during the first meeting. Minutes from this CALMAC meeting, along with the material used at the meeting, were developed, edited for completeness, and placed on the CALMAC web site for public consultation.

Following this meeting the draft work plan was finalized based on the comments and feedback received from the Advisory Group and from the public. A final Work Plan was then submitted for review to the Project Advisory Group. Comments were incorporated to produce a final Work Plan that was used to guide the project.

The Project Advisory Group provided general and topic-specific project guidance and reviewed all chapters of the Framework as they were developed. The Advisory Group also provided consistent, ongoing guidance to the effort and provided valuable feedback as the chapters were developed. However, that does not mean that a consensus was achieved on all decisions regarding the Framework effort. Several rounds of Advisory Group and public input were provided in order for the scope and focus of the Framework to be finalized. Nevertheless, this document is the sole responsibility of the TecMarket Works Team that produced it. (Therefore, responsibility for the contents of this document needs to be borne most heavily by the leaders of this team.)

The Project Advisory Group members and their affiliations are:

- Marian Brown, Project Manager, Advisory Group Chairperson, Southern California Edison
- Eli Kollman and Jay Luboff, Energy Division, California Public Utilities Commission

- Don Schultz and Christine Tam, Office of Ratepayer Advocates, California Public Utilities Commission
- Mike Messenger and Sylvia Bender, California Energy Commission
- Valerie Richardson, Chris Ann Dickerson, and Kenneth James, Pacific Gas and Electric Company
- Athena Besa and Rob Rubin, Sempra Energy

The TecMarket Works Framework Team consists of:

- Nick Hall, TecMarket Works
- Lori Megdal, Megdal & Associates
- Pete Jacobs and Stuart Waterbury, Architectural Energy Corporation
- Roger Wright, RLW Analytics
- Paul Chernick, Resource Insight Incorporated
- Ken Keating, Ken Keating and Associates
- Sharyn Barata, B&B Resources
- Ed Vine, Ed Vine and Associates
- Steve Nadel and Marty Kushler, American Council for an Energy Efficient Economy
- Ralph Prah, Ralph Prah and Associates
- John Reed, Innovologie

The TecMarket Works Framework Team was organized into several expert teams based on the specific evaluation type and development work areas. These teams are shown in Figure 2.1.

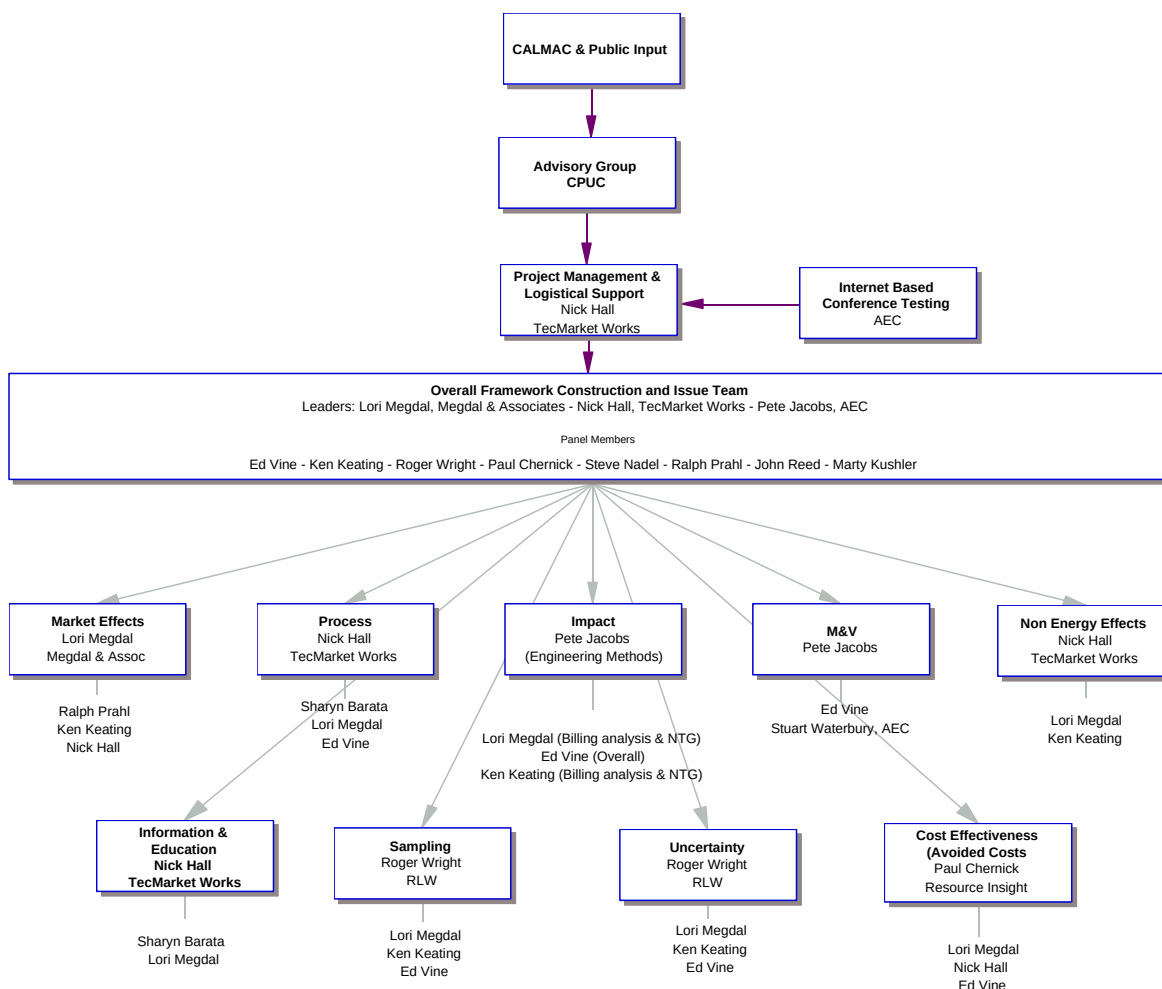


Figure 2.1: Project Team Assignments and Organization

Literature Review

One of the first project tasks was a literature review to support the development of the Framework. The primary purpose of the literature review was to not have the Framework re-invent the evaluation wheel in California, but to use currently available literature to guide the Framework development process.

The individual chapters of the Framework correspond to specific types of evaluation research and provide valuable information on determining which evaluations to conduct for the different types of energy efficiency and procurement programs offered in California. These chapters focus on the following type of research:

- Impact Evaluation,
- Measurement and Verification (M&V),
- Process Evaluation,
- Information and Education Program Evaluation,
- Market Transformation (MT) Program Evaluation, and
- Non-Energy Effects (NEE) Evaluation.

In addition to these evaluation-focused chapters there are two additional chapters that cut across all of these types of evaluation that are incorporated into the Framework; these are chapters on uncertainty in the evaluation process, and sampling methods and approaches.

The chapters on the specific types of evaluation provide an overview of the evaluation methods and issues to be addressed for each type of evaluation, some indications of what defines a high quality evaluation for each type of research effort, and the presentation of key references for additional technical background and to provide examples of use. One of the key documents that heavily informed certain chapters of this Framework is the prior California Measurement & Evaluation Protocols.⁵ Other key reference documents include the Energy Efficiency Policy Manual⁶ and the International Performance Measurement and Verification Protocol.⁷

Other “protocol” type documents used for general reference in the early stages of this project included the following publications:

- USEPA (1995). *Conservation Verification Protocols: A Guidance Document for Electric Utilities Affected by the Acid Rain Program*.
- * FEMP (2000). *Federal Energy Management Program (FEMP) M&V Guidelines: Measurement and Verification for Federal Energy Projects*. Federal Energy Management Program. September. Version 2.2, DOE/GO-102000-0960.
- * ASHRAE (2002). *Measurement of Energy and Demand Savings, Guideline 14*. American Society of Heating, Refrigeration and Air Conditioning Engineers: Atlanta, GA. Available at <www.ASHRAE.org>.
- Nexant and Lawrence Berkeley National Laboratory (2002). *Detailed Guidelines for FEMP M&V Option A*. Federal Energy Management Program. Available at <<http://ateam.lbl.gov/mv>>.
- AIS, SRC International (2001). *European Ex-post Evaluation Guidebook for DSM and EE Services Programmes*. International Energy Agency. April. Available at: <http://dsm.iea.org/> Report is available under the library.
- Xenergy, ADM Associates, VACom Technologies and Partnership for Resource Conservation (2001). *2001 DEER (Database for Energy Efficiency Resources) Update Study*. California Energy Commission. Study ID 3001. Within searchable database at <<http://www.calmac.org>>.
- * Violette, Daniel (1995). *Evaluation, Verification, and Performance Measurement of Energy Efficiency Programs*. International Energy Agency. Available at <<http://dsm.iea.org/>>. Report is available under the library.

⁵ *Protocols and Procedures for the Verification of Costs, Benefits, and Shareholder Earnings from Demand-Side Management Programs*. (* CADMAC 1999). Referenced as the Measurement and Evaluation Protocols. Available at: www.calmac.org.

⁶ *Energy Efficiency Policy Manual, Version 2*. August. (* CPUC 2003). Available at <<http://www.cpuc.ca.gov>>.

⁷ *IPMVP Volume 1: Concepts and Options for Determining Energy and Water Savings*. (International Performance Measurement and Verification Protocol 2001). Available at <<http://www.ipmvp.org/download.html>>.

A focus on how the literature would be used for the Framework as a whole and for individual types of evaluation guided the selection of documents for the literature review. The criteria for selecting literature for review from the thousands of available publications centered upon providing support for:

- Defining types of evaluations needed for various decisions being made;
- Criteria for prioritizing between evaluation needs;
- Graduate textbooks and seminal works in fields related to energy efficiency program evaluation, to include:
 - o Research design
 - o Sampling design
 - o Survey design
 - o Statistics and econometrics
 - o Evaluation studies
 - o Program theory/logic work and evaluation
 - o Energy engineering;
- Work that provided testing or advancements of new methods, tests, and specific examples in the following types of efficiency evaluation:
 - o Energy impact and program effects evaluation
 - o Measurement & verification
 - o Process evaluation
 - o Market assessment/progress measurement
 - o Evaluating training efforts;
- Advancements and tests within energy efficiency evaluation in the areas of:
 - o Uncertainty
 - o Sampling;
- Other aspects of an overall evaluation research as were supportive in the creation of the integrated evaluation planning, evaluation, and program planning components within the umbrella roadmap.

All key team members, several Advisory Group members, and a number of public commenters provided recommendations for the literature to be reviewed for this effort. From these recommendations a literature review list was developed for possible use in the project. This list was reviewed by team members and then by the Project Advisory Group and finalized for the project effort. This list is provided in Appendix B, the Bibliography. Those items in the bibliography (and any other areas of the Framework with citations) that were part of the initial literature review list are indicated by asterisks in the front of the citation (*).

Document accessibility was an important criterion for whether a possible reference became a primary information source for the project. If documents were readily available for free, were made available through the CALMAC web site, could be made available through the project web site <<http://www.tecmarket.net/CAFrame.htm>>, or were available at reasonable costs (such as various conference proceedings), they could serve as a primary project reference. Expensive documents or documents of limited availability, (e.g., EPRI documents) may be referenced in the Framework, but are not

emphasized as primary reference documents. Similarly, older documents that might be difficult to access are generally not primary references for the project, but may be referenced.

Development of Principles and Roadmap

The next step in the Framework construction process was to develop guiding principles for each type of evaluation. These were the primary paradigm parameters set for the Team and Advisory Group evaluation discussions. Frequent teleconferences with the Project Advisory Group were held to discuss evaluation issues, concepts, and the design and construction of the draft roadmaps. This process provided significant input and discussion between the Project Advisory Group and the key project leads to inform the development of the basics of the Framework and its roadmap characteristics. During this process a discussion packet of material was developed that contained these principles and draft roadmaps for each type of evaluation, the umbrella framework, and the overarching evaluation studies and issues. This packet was also distributed for public review to self-selected interest groups (see the public input process discussion below). The material was then revised as appropriate based upon the public input and used to guide the development of the roadmaps for public presentation at a set of CALMAC workshops.

An assembly of this material was then prepared for public review and comment during an open CALMAC meeting and placed on the CALMAC web site prior to the meeting. During the public meeting the material was presented and discussed, and public comments and recommendations on the material were received by the project team. The CALMAC meetings were held in southern California (on September 3, 2003) and northern California (on September 4, 2003). Following these meetings, minutes were prepared and placed on the CALMAC web site for public consultation.

Development of Draft and Final Project Reports

All of the material developed in these early efforts, the comments received at the CALMAC meetings, and the literature review provided the raw material from which the project teams developed the draft roadmaps and chapters for the Framework document. During the development of the Framework chapters a significant effort was made to ensure that references were incorporated into the chapter text to provide appropriate citations to support the chapter text but also to site reference material that could be used to learn more about the methods, tools, and issues being presented. This included citing references to academic literature upon which the foundations of the methods for quality evaluations are based, and the energy efficiency evaluation literature to provide examples of the use of the methods, issues, and tools discussed. A complete bibliography is included as Appendix B.

An initial draft document was distributed to the Project Advisory Group and to the public interest group list in November 2003. Comments were received and reviewed by the TecMarket Works Framework Team in December 2003. From this review, a list of areas of disagreement, conflicting recommendations, and important direction issues was

compiled. This list, with an outline of the areas needing decisions, was provided to the Project Advisory Group. Lengthy teleconferences with the Project Advisory Group were held on December 23rd and 29th of 2003 to resolve these issues and provide final document guidance.

A complete set of revised chapters was prepared for a second round of review by the Project Advisory Group. These were sent out for review at the end of January 2004.

Comments from the Project Advisory Group on the second draft of the Framework document were used to create the final draft. The final draft document was posted on the CALMAC web site in mid February for public review. A notice of this posting with a request for public feedback was sent to the CALMAC service list at the time of the posting. Public feedback was accepted by e-mail and used to guide the development of the final Framework document.

Public Input Process

This project has one of the highest levels of public input experienced within an evaluation project. First, there was an open project kick-off meeting that discussed the details concerning possible subjects and issues to be addressed in the projects. Over 150 comments were received during this meeting and incorporated into a publicly available set of meeting presentation materials and meeting notes posted on the CALMAC web site.

Then midway through the project two open CALMAC meetings were held. One was held in northern California and the other was in southern California in early September 2003. These meetings presented the evaluation guiding principles that were proposed for each evaluation type, and the issues to be presented in the Framework document. Comments and discussion were received at both meetings. Publicly available meeting presentation materials and notes were prepared and posted on the CALMAC web site.

Additionally, this project developed a unique public input process based upon self-selected subject interest group lists. To construct these lists an e-mail was sent to the CALMAC service list asking that anyone interested in reviewing the Framework's progress and draft chapters as they developed could be sent this material if they responded by e-mail asking to be placed on one or more of the chapter review lists. The evaluation principles and the draft Framework were sent to the interest group lists. In addition, the draft materials for the September CALMAC meeting were sent to this group. Then again, as the Framework draft chapters were developed in November, they were each e-mailed to members on the interest group list for that section of the Framework.

After multiple revisions to incorporate public comments, a final draft Framework was provided for public input. This document was placed on the CALMAC web site and

announced for public review and comments. This public feedback process provided the final input opportunity for the project just prior to finalizing the Framework document.

Chapter 3: Use of the Evaluation Framework by Different Stakeholders

One of the early discussions between the California Public Utility Commission (CPUC) staff, the Framework's Advisory Group, and the TecMarket Works project management team involved identifying for whom the Framework document should be written and the level of detail to be included in the various chapters within the Framework document. As a result of these discussions, this document is written so that readers with a limited background in the field of evaluation can understand the scope and application of the evaluation efforts and concepts presented. At the same time, the Framework is meant to be a reference document for evaluators and a tool to help design high quality evaluations. These two conditions required a balance in the material presented that attempts to meet the information needs of both evaluation and non-evaluation professionals.

The document also seeks to provide enough information and references so that evaluators will know what methods and levels of rigor are expected. To this end, the document includes a significant list of references guiding readers to additional information about the evaluation methods, tools, or issues being discussed. It is also realized that different chapters need to be produced at different levels of specificity. For example, the impact evaluation chapter is written at a higher level of detail than other chapters because specific impact evaluation and data handling practices need to be identified and discussed to indicate their strengths and weaknesses for use within the impact evaluation process.

In identifying the focus and scope of the Framework effort the CPUC, the Project Advisory Group, and the project team stipulates that professionals specifically trained and practiced in the type of evaluation for which they are responsible, should conduct those evaluations. These parties accept this stipulation because it is imperative that people who are experts at assessing energy programs conduct evaluation studies. The Framework also stipulates that program evaluations will be conducted by firms, organizations, or groups that are independent of the implementation administrator or contractor and that the evaluation teams will maintain an arm's-length relationship with implementation administrators and contractors in order to help assure objective and reliable evaluation efforts. This document assumes that skilled professionals, who do not require that the Framework document be a detailed instruction manual, will conduct evaluations, and that the Framework document be designed to focus and shepherd the evaluation process. This focusing and shepherding is intended to guide evaluations toward producing rigorous, reliable energy program evaluation results that support fair comparisons across programs to the extent possible within the available evaluation resources.

The Framework document is written with the following types of individuals in mind.

- Regulatory staff and policy makers responsible for policy-associated goal setting and resulting high-level program or program portfolio decisions.
- Portfolio managers responsible for guiding or developing program portfolios.

- Power planning personnel responsible for power supply acquisition.
- Evaluation oversight managers and reviewers responsible for guiding or overseeing the evaluation efforts or interpreting evaluation results.
- Program administrators and implementers responsible for evaluation support efforts for their programs.
- Cost-effectiveness and avoided cost personnel responsible for identifying programs or program components that provide cost-effective energy supplies.
- Evaluation designers and managers responsible for designing or recommending evaluation methods and approaches that provide reliable results.
- Impact evaluators responsible for conducting studies which estimate the energy impacts (kW, kWh, therms, etc.) associated with a program, a group of programs, or a program portfolio.
- Metering and monitoring evaluators responsible for designing or conducting field efforts to support the impact evaluation efforts.
- Process evaluators who are responsible for designing and conducting process evaluations to improve operations and program cost-effectiveness.
- Information and education evaluators responsible for designing and conducting evaluations of information and educational programs to assess program effects.
- Market transformation program and market effects evaluators responsible for designing and conducting evaluations of program markets, including market conditions, market operations, market effects, or market baseline studies.
- Non-energy effects evaluators who are responsible for designing and conducting studies of the non-energy effects of programs, groups of programs, or program portfolios.
- Statisticians, research data managers, and analysts who are responsible for data handling and analysis efforts.
- Other stakeholders and interested parties that have an interest in the evaluation of energy programs or in the results of these studies.

Developing a single document for such a wide diversity of interested stakeholders is a challenge in itself. The Framework team attempts to meet this challenge by providing a document that does not over-simplify the evaluation efforts, and at the same time does not delve into the details of conducting program evaluations.

With these conditions in mind the Framework team realizes that specific parts of the Framework document will be of greater or less interest to the types of professionals identified above. The authors have developed two mechanisms to aid the reading and use of this document by different stakeholders. First, many of the Framework's chapters contain a Preface describing the contents of the chapter and indicating the various stakeholders that might find value from that chapter. Second, the authors have developed

a summary table (below) indicating which of the types of individuals identified above might find value in each of the Framework’s chapters.

The following table presents the specific chapters within the Framework document that the authors suggest will be of interest to each of the categories of professionals identified above. Across the top of the table are the types of professionals identified above. Within each square of the table matrix is a ball indicating the level of interest that the authors suggest is appropriate for each of the types of professionals identified.

There are three levels of “interest categories” in this table consisting of a black bullet (●), a small bullet in a circle (⊙), and a white bullet (○). The interest category populated by a black bullet (●) indicates that the authors suggest the chapter be considered important information for the types of professionals identified. Professionals in the categories that have a black bullet in the cell corresponding to the chapter on the left should, in the opinion of the authors, read these chapters carefully and have a working knowledge of the information within those chapters. Professionals with a small bullet in a circle (⊙) in the chapter cell may want to consider the material in the chapter as suggested reading, and be familiar with the chapter contents and implications for the evaluation effort. Finally, the professional categories that have a white bullet in the chapter cell (○) might want to have at least an overview familiarity with the chapter contents.

By reviewing the table and examining the bullets within the matrices, readers can focus their document review efforts on the chapters of the Framework that are most appropriate to their areas of responsibility.

Table 3.1: Suggested Framework Reading by Various Stakeholders

● Important information ⊙ Suggested reading ○ Some familiarity	Regulatory Staff & Policy Makers	Portfolio Managers	Power Planning Personnel	Evaluation Oversight Managers & Reviewers	Program Administrators / Program Implementers	Cost-Effectiveness and Avoided Cost Personnel	Evaluation Designers & Managers	Impact Evaluators*	Metering & Monitoring Evaluators	Process Evaluators	Information & Education Evaluators	Market Transformation and Market Effects Evaluators	Non-Energy Effects Evaluators	Statisticians & Research Data Managers	Other Stakeholders & Interested Parties
1. Introduction and Purpose of Framework	⊙	⊙	⊙	●	●	⊙	●	●	●	●	●	●	●	○	⊙
2. Study Methodology and Description of Public Input Process	⊙	○	○	○	⊙	○	⊙	⊙	⊙	⊙	⊙	⊙	⊙	○	○
3. Use of the Framework by Different Stakeholders	●	⊙	⊙	●	●	⊙	●	●	●	●	●	●	●	⊙	⊙
4. Evaluation Overview and Issues	●	●	○	●	●	●	●	●	●	●	●	●	●	⊙	⊙
5. Umbrella Roadmap for Evaluation Planning and Prioritization Decisions	●	⊙	⊙	●	●	●	●	●	●	●	●	●	●	●	⊙
6. Impact Evaluation Roadmap	⊙	⊙	⊙	●	●	⊙	●	●	⊙	○			○	⊙	○
7. Measurement and Verification Roadmap	⊙	⊙	⊙	●	●	○	●	●	●	○				⊙	○
8. Process Evaluation Roadmap	⊙	○		●	●	○	●	⊙	⊙	●	●	●	⊙	⊙	○
9. Information/Education Roadmap	⊙	⊙		●	●	○	●	○		○	●	○	○	⊙	○

Table 3.1: Continued

● Important information ⊙ Suggested reading ○ Some familiarity	Regulatory Staff & Policy Makers	Portfolio Managers	Power Planning Personnel	Evaluation Oversight Managers & Reviewers	Program Administrators / Program Implementers	Cost-Effectiveness and Avoided Cost Personnel	Evaluation Designers & Managers	Impact Evaluators*	Metering & Monitoring Evaluators	Process Evaluators	Information & Education Evaluators	Market and Market Effects Evaluators	Non-Energy Effects Evaluators	Statisticians & Research Data Managers	Other Stakeholders & Interested Parties
10. Market Transformation Program Evaluation Roadmap	⊙	⊙	⊙	●	●		●	⊙	○	⊙	⊙	●	⊙	⊙	○
11. Non-Energy Effects Evaluation Roadmap	⊙	○	○	●	●	⊙	●	⊙	○	⊙	⊙	●	●	○	○
12. Uncertainty	●	●	●	●	⊙	⊙	●	●	●	●	●	●	●	●	⊙
13. Sampling	⊙	○	○	●	⊙		●	●	●	●	●	●	●	●	○
14. Evaluation and Cost-Effectiveness	●	●	⊙	●	●	●	●	●	●	●	●	●	●	●	⊙
15. Overarching Evaluation Studies	●	●	⊙	●	●	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	○	○
Appendix A: Glossary of Terms	⊙	⊙	⊙	●	●	●	●	●	●	●	●	●	●	●	⊙
Appendix B: Bibliography	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Appendix C: Guidelines for Evaluation Planning	⊙	⊙	○	●	●	⊙	●	●	●	●	●	●	●	●	○

* Sections of the Impact Chapter may contain information that the authors consider critical knowledge for engineering and statistical billing evaluators, depending upon the analysis method selected.

Chapter 4: Evaluation Overview and Issues

Preface

This chapter of the Framework presents and discusses an overview of energy program evaluation, the different types of evaluation, the history of energy efficiency program evaluation in California, the use of program theory within evaluation, evaluation ethics, and the role of evaluation in the formation of energy policy. This chapter serves as a foundation for understanding the Framework and provides information for the development of evaluation policy in California.

Due to the groundwork this chapter provides and the importance of the topics covered with regard to the development and implementation of a statewide evaluation Framework, a wide range of evaluation stakeholders can benefit from this chapter, including:

- Regulatory Staff & Policy Makers;
- Evaluation Oversight Managers & Reviewers;
- Program Administrators/Program Implementers;
- Cost-Effectiveness and Avoided Cost Personnel;
- Evaluation Designers & Managers;
- Evaluators for all types of evaluation - impact, metering and monitoring, process, information and education programs, market transformation and market effects, and non-energy effects;
- Statisticians & Research Data Managers; and
- Portfolio Managers.

Core Purpose of Evaluation

When distilled to its most basic level, the essential over-arching purpose of evaluation is to help ensure that good decisions are made regarding the investment of energy program resources by providing rigorous, independent evaluation studies and study results. While the evaluation purpose is pursued through many different types of activities and methodologies by different parties and in different timeframes across the life cycle of a program, it is useful to keep this over-arching purpose in mind as a unifying theme for why evaluations are funded and conducted.

One of the primary ways in which evaluations provide information for making good decisions is testing the implicit and explicit assumptions within the program's program theory. It may also be important for the evaluation to develop and test alternative theories or hypotheses that may compete with or supplement the program theory. For

example, important conditions to be tested by the impact evaluation are the energy and demand savings per measure, per customer, and for the program overall - and the reason for those savings. The process evaluation may test whether the program is being implemented as planned, the various reasons driving customer participation and satisfaction, the usefulness of tracking data, etc. The market evaluation can test the assumptions held within the program's logic model and alternative hypotheses of how the market operates and how, and whether, the program interventions are obtaining the anticipated outputs, short-term outcomes, intermediate outcomes, long-term outcomes, and ultimate outcomes.

The Two Key Functions of Evaluation – Summative and Formative Evaluations

There are two key functions that are at the core of the evaluation enterprise:

1. To document and measure the effects of a program, and
2. To help understand why those effects occurred and identify ways to improve the program.

Most parties would agree that the single most important function of evaluation is to document and measure program effects. Indeed, determining the impacts of a program or group of programs is the threshold requirement for being able to assure accountability for the expenditure of resources on that program. These are summative evaluations.⁸ Summative evaluations are done after the program has been operating to document program impacts and are used to inform decisions on whether to continue, expand, cut back, or end the program. Generally, impact evaluations in the energy efficiency field are summative evaluations.

However, the second key function (understanding why the observed effects occurred and identifying ways to improve program effectiveness) is also important. This function is supported through formative evaluations. Formative evaluations are often provided either early in the program's operations or steady-state programs to obtain feedback and discover ways to improve a program. Process evaluations are typically used as formative evaluations.

The importance of this perspective can be seen through a simple example. Consider the situation where the observed impacts of a program are weak, but this effect is due to an easily correctible problem in the program implementation. Simply reacting to the impact evaluation results alone might well lead to a technically accurate but policy-poor decision to terminate the program, whereas understanding why those results occurred and what

⁸ The terms and distinctions identified as summative and formative evaluations have become uniformly adopted. They were first coined by Scriven in 1967 with regard to educational evaluation (Scriven 1967). Further discussions and comparisons can be found in: *Program Evaluation: Alternative Approaches and Practical Guidelines*, 2nd edition (Worthen et al. 1997), pp. 14-18; and *Evaluation: Methods for Studying Programs and Policies* (* Weiss 1998), pp. 31 -33.

can be done to improve program results can lead to improvements in the program implementation and ultimately a successful program.

Similarly, experts realize that new programs typically have to experience a “start-up” phase, where effectiveness is not yet optimized and initial costs are higher. It would be inappropriate to simply look at initial impact results and make a quick judgment about a program’s potential without understanding where the program was in terms of its life cycle.

The new California Evaluation Framework is built with the understanding that the two key concepts, elements of summative and formative evaluations, are important to the overall policy decisions that would be made based upon evaluation results. High quality evaluation efforts would incorporate both concepts into their energy program evaluation design: (1) the evaluation would document program impacts and (2) the evaluation would provide for a better understanding of the observed results, the reasons for the observed results, and identify applicable opportunities for improving program performance.

These principles translate well to the California context. In California, it is incontrovertible that the primary purpose of evaluation is to document the amount of ex-post or net energy saved through the Public Goods Charge programs and to provide information to help determine the cost-effectiveness of acquiring those resources. However, it is also important that evaluations provide information to help improve programs and assist in making the best possible choices in a public policy context. California is known for incorporating multiple and complex objectives in its public policy directives. This makes it ever more important to utilize as comprehensive and rigorous evaluation approach as cost-effectively as possible in order to provide policy makers and regulators the information they need to make sound decisions.

The Two Basic Types of Evaluation

Proceeding to the final level of this conceptual framework, there are two basic types of evaluation that are used in energy efficiency evaluation to support the two key functions and the fundamental over-arching purpose of evaluation. These two broad categories of evaluation are:

1. Effects Evaluation
2. Process Evaluation

As the names imply, these evaluation types correspond directly to the two key evaluation functions outlined earlier. Effects evaluation includes many different types of program “effects.” These include the energy and demand savings that are the focus of impact evaluations. Impact evaluation comes in many forms, ranging from simple engineering estimates applied to counts of measures installed to sophisticated statistical models incorporating measured energy consumption data and numerous other variables and statistical corrections. Other types of programs can have different types of effects

evaluations, such as market effects and transformation studies for market transformation programs, and information and education program effects evaluation. Effects evaluation also increasingly has come to include important elements that go beyond the effects that are the primary program goals to secondary effects to “non-energy effects.” The one unifying theme in all these activities is the fundamental purpose of effects evaluation: to serve the key function of documenting and measuring effects caused by the program.

Similarly, process evaluation can be accomplished using a variety of methodological techniques, but its purpose is to serve the key function of documenting and understanding how the program is operating and identifying opportunities for program improvement.

An expanded view of process evaluation can include studies to understand market operations and processes and how market transformation programs are operating within those market processes. In this way, the process and effects components of evaluation can work together to provide both the functions of effects and process evaluation for market transformation programs.

It is possible to further delineate several sub-categories of market transformation program evaluation, each with a somewhat different focus and purpose. In this regard, the Framework would distinguish:

1. Market effects evaluations, which examine the changes in a market caused by a program,
2. Market operations and baseline evaluation,⁹ which looks at how a market operates, the key information points, information hubs, and how products flow. This type of evaluation is used by evaluators for assessing program performance or program designs and operations,
3. Marketing evaluations, which examine the effectiveness of a set of marketing, or market, outreach efforts.

In many cases a well-designed and comprehensive evaluation approach will incorporate both of the basic types of evaluation: effects and process to provide a complete assessment of the program so that program and policy decisions can be grounded in the rigorous and objective information needed to support a program or policy decision.

Program Theory/Logic Model Use in Evaluation

An important component of the evaluation effort is to draw upon the program theory and logic model, to include its review (or development if one is not available) and use as an evaluation planning tool. This section presents an introduction to program theory and logic models and their use in evaluation. It also provides references that readers unfamiliar with the program theory literature can use to learn about its uses, methods, and importance.

⁹ Sometimes referred to as “market characterization.”

A program theory essentially tells the story of a program, detailing what a program is going to do, why it is going to do it, and what will be accomplished as a result.

A program theory is a presentation of the goals of a program, incorporated with a detailed presentation of the activities that the program will use to accomplish those goals and the identification of the causal relationships between the activities and the program's effects. The program theory describes, in detail, the expected causal relationships between program goals and program activities in a way that allows the reader to understand why the proposed program activities are expected to result in the accomplishment of the program goals. A well-developed program theory can (and should) also describe the barriers that will be overcome in order to accomplish the goals and clearly describe how the program activities are expected to overcome those barriers. A program theory may also indicate (from the developers perspective) what program progress and goal attainment metrics should be tracked in order to assess program effects.

Program theories (PT) are sometimes called the program logic model (LM). A stricter definition would be to differentiate the program theory as the textual description while the logic model is the graphical representation of the program theory showing the flow between activities, their outputs, and subsequent short-term, intermediate, and long-term outcomes. Often the logic model is displayed with these elements in boxes and the causal flow being shown by arrows from one to the others in the program logic. It can also be displayed as a table with the linear relationship presented by the rows in the table. The interactions between activities, outputs, and outcomes are critical to understanding the program logic and argue for the need to have, or construct, both a program theory and a program logic model. An example of a logic model diagram is provided in the Appendix to this chapter.

The full range of known external influences is also important to display in the logic model. Without them, these influences may be forgotten or underestimated, and an inaccurate evaluation can result from failing to understand the context in which the hypotheses and causal mechanisms are to be tested.

The development and refinement of program theories and logic models and their use within evaluation has occurred in the larger evaluation field outside of energy program evaluation. There is a rich history and experience in using these tools for better program design, program management, and evaluation in education, public health, and social programs. It is important for evaluators to refer to the program theory literature as they assess program-developed PT/LM or when they develop alternative PT/LM in order to plan their evaluations, unless they have extensive experience in this area. Textbooks providing a good background and foundation for this work can be found in works by Weiss,¹⁰ Rogers et al.,¹¹ and McLaughlin and Jordan.^{12, 13} Background information was also provided in the *2001 Framework Study* that is well worth examining.¹⁴

¹⁰ *Evaluation: Methods for Studying Programs and Policies* (* Weiss 1998).

¹¹ *Program Theory in Evaluation: Challenges and Opportunities*. (* Rogers et al. 2000b).

It is equally important to keep an eye on the developing state-of-art of using PT/LM in the energy evaluation field as significant developments continue to occur. Early articles using PT/LM within energy program evaluation can be found in Jordan et al.¹⁵ and Ives-Petersen.¹⁶ A program theory evaluation approach was used to expand the market barriers perspective to include communications theory and diffusion factors in a commercial program evaluation in California.¹⁷ PT/LM have been very helpful in identifying indicators for market transformation (MT) programs,¹⁸ and to align indicators with performance metrics for MT programs.¹⁹ It is also a critical component in developing evidence of causality for MT efforts.²⁰ At the same time, work in efficiency evaluation has noted the importance that PT/LM work be appropriately grounded in market theory and recognize the business paradigms of the program interventions being implemented.^{21, 22} The PT/LM effort should also, where possible, be expanded to reference the market research that was used to identify the causal relationships between the program activities and the program goals.

A potentially useful program analysis can be an assessment of the program's PT/LM against the social science, marketing, and communication theories that relate to the behaviors or decisions the program is attempting to change. These other fields of study can sometimes help provide insight into the likelihood of the program's actions resulting in a desired program effect. Similarly, evaluations of other programs can often inform an assessment of the PT/LM and the ability of the actions and relationships presented in the PT/LM to have the intended effect within the target market.

The PT/LM approach can be equally important in evaluating educational efforts.²³ In fact, the importance of the use of PT/LM within MT program evaluation and in evaluating information and educational programs is why assessing and using PT/LM are key elements within the chapters on these types of evaluations in this Framework.

-
- ¹² "Logic Models: A Tool for Describing Program Theory and Performance." (McLaughlin and Jordan 2004).
- ¹³ "Logic Models: A Tool for Telling Your Performance Story." (* McLaughlin and Jordon 1999).
- ¹⁴ *A Framework for Planning and Assessing Publicly Funded Energy Efficiency.* (* Sebold et al. 2001).
- ¹⁵ "Measuring and Managing the Performance of Energy Programs: A Case Study." (Jordan et al. 1997).
- ¹⁶ "Using the Program Logic Model to Increase the Relevance and Use of Evaluation Findings of Market Transformation Projects." (* Ives-Petersen 1999).
- ¹⁷ "Using Diffusion and Communications Theory to Expand Market Barrier Examination in MT Measurement." (Megdal et al. 1999).
- ¹⁸ "A Systematic Application of Theory-Based Implementation and Evaluation of Market Transformation Programs." (* Hastie et al. 2000).
- ¹⁹ "Using Program Theories to Align Performance Metrics with Public Purpose Goals." (Erickson and Bloch 2002).
- ²⁰ "He Did It! He Did It! – Providing Evidence for Causality." (* Megdal et al. 2001).
- ²¹ "Merging Program Theory and Market Theory in the Evaluation Planning Process." (* Hall and Reed 2001).
- ²² "When Business Analysis Tools Need to Accompany Program Theory Evaluation within Energy Efficiency Market Transformation Efforts." (Megdal et al. 2001).
- ²³ "Detecting Behavioral Change from a Visit to a Children's Museum Energy Conservation Exhibit." (* Peters et al. 2000).

By spelling out the theory, the program evaluator will be able to identify and examine the assumptions underlying the theory by comparing the theory to what happens as the program is implemented. If the assumptions are supported, then there is reason to believe that the program is working as planned and for the reasons indicated in the program theory. If the assumptions are not supported or other alternative assumptions are identified, then the program may not be working, or not working as efficiently as it could. In these cases the program theory needs to be modified to better reflect the actual operation of the market so that the evaluation can be on the program activities that are expected to provide the greatest benefits to the program. However, if the program theory or the assumptions about the market are not accurate, the program may need to be redesigned to increase the effects from the program. Used properly, the program theory can be used to help identify the issues to address in the evaluation, and to identify activities of the program that need to be re-examined.

An evaluation design and program theory can work hand-in-hand to be able to have the evaluation differentiate between theory failure (incomplete or inaccurate theory), and program failure (poorly designed or implemented operational procedures). In this case of theory failure, the program theory is incorrect or not complete. The assumptions about the market or the causal mechanisms that create attitudes or behaviors are not valid, are only partially valid, or are missing key theoretical components. The program theory needs to change and the program needs to be refined accordingly.

On the other hand, if the evaluation identifies program failure, this means that the theory appears to be correct. However, the program implementation had problems that did not allow it to have the anticipated outputs and initial subsequent outcomes. Figure 4.1 shows the differences between theory failure and program failure.

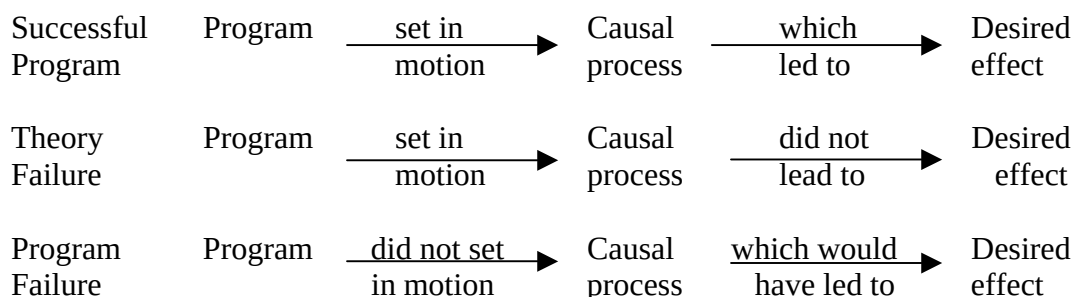


Figure 4.1: Theory Failure & Program Implementation Failure²⁴

Developing a Program Theory

The easiest way to develop a program theory is to start by systematically describing a program in terms of resources, activities, outputs, short-term outcomes, and long-term outcomes. (See the literature cited above in the first subsection of the PT/LM section.)

²⁴ *Evaluation: Methods for Studying Programs and Policies.* (* Weiss 1998), page 129.

- *Inputs (resources)* are the elements required by an organization, program, or project to initiate and/or sustain activities. Examples are money, collaborations, skills, and time.
- *Activities* are the program activities that are used to produce the outputs that initiate the causal logic within the logic model.
- *Market actors* are those market actors targeted by the interventions or that play a role in the causal logic of the program theory.
- *Outputs* are the immediate results of an activity. Examples are the number of contacts made, number of brochures printed, number of contractors recruited, and number of audits completed.
- *Outcomes* are the intermediate or once removed consequences resulting from program activities and program outputs. There may be a sequence of outcomes. Outcomes may be unintended or intended but they are not prompted by direct action on the part of the program. Examples are changes in awareness, attitudes, and behaviors, participants referring non-participants to the program, trade publications running articles about efficient equipment and practices, dealers changing their stocking practices, etc.
- *Long-term outcomes (impacts)* are the end-states to be realized. Impacts may take months or years to accomplish and may be influenced indirectly by the intervener's actions. Impacts are the long-term goals of the program. Examples are kWh saved, gallons of water saved, tons of CO₂ reduced, efficient technologies and practices are the industry standard, T-12 fluorescents are difficult to buy, etc.

Once a program has been systematically described in terms of resources, outputs, outcomes, and long-term impacts, the procedure for developing a program theory is a systematic one. One of the best ways to develop a program theory is to start with the long-term outcomes and work backwards to resources. Essentially, the process is one of repeatedly asking the same question, if “Z” is a long-term outcome (or short-term outcome, output, or activity), what is required to produce “Z.” It is then a matter of writing the causal relation in the form of a statement: “Y” will cause “Z.” One then backs up and asks what will cause Y and continues until one has described the required activities and resources.

One can then reverse the order and edit the statements until one has a sequence of causal statements that describe how the program works.

Several things are likely to happen as this is being worked upon. It is likely that gaps in the causal relationships between actions and expected effects will be found. Some steps will be identified that require substantial leaps that suggest that the theory needs further refinement. Some steps in the theory will seem quite improbable, suggesting that the theory, and probably the program design, needs improvement. Some steps will contradict what is known from the program, marketing, and evaluation literature and other social science and business theories. Sometimes an assumption, key to the way the program is designed for one of its causal chains, can be in conflict with an assumption in a different

causal chain. All of these instances identify places where the program theory and the program design may need improvement.

This process and the resulting program theory and logic model provides program information that helps isolate important measurement indicators. These are indicators for the outputs and outcomes of the program. More importantly, the program has been described in a way that will enable the evaluator to identify key issues and to identify important questions on program assumptions or key hypotheses that the evaluation needs to be able to answer.

An example of a Program Theory is provided in the Appendix to this Chapter.

Reasons for an Evaluator to Use and Refine the Program Theory

Ideally, the program evaluator should not find it necessary to develop a program theory to support the evaluation planning process. All programs should have a detailed program theory that is used to justify the need for the program, justify the activities needed to accomplish the program goals, and determine the funding levels needed to successfully implement the program. In the past, program theories have often been embedded in the general logic associated with a program's efforts and not specifically developed into a formal program theory. However, in the last several years organizations that provide funding for energy efficiency programs have begun to require detailed program theories so that they can make program and program funding decisions to help assure that public benefits dollars are well spent and that there is a high probability that program goals will be obtained. When a program theory is already developed the evaluation planning efforts should be informed by that "official" program theory. However, evaluators should not rely only on the official program theory for their evaluation planning efforts. When evaluators examine the official program theory it is not unusual for the evaluator to identify alternative paths not reflected in the official program theory by which participants can reach the same desired outcomes as those reflected in the official program theory. However, when a program does not have a formal program theory the evaluation team should develop a program theory in cooperation with the program administrators that can be used to inform the evaluation planning process. Regardless of the presence or absence of an official program-developed program theory, it is up to the evaluator to identify research approaches that test for the different causes for the effects that are projected by the program theory.

There are a variety of reasons why a program theory should be used for developing an evaluation plan. These include:

- The program theory can be used to examine and clarify program assumptions and program goals so that the evaluation team has a complete understanding of the program to be evaluated.
- The program theory can be used to help identify potentially missing or extraneous activities that need to be addressed in the evaluation.

- The program theory provides a program description that can be used to:
 - Identify key hypotheses to be tested as part of the evaluation.
 - Identify key measures for which data need to be collected.
- Because the program theory provides a logic and sequence, it can be used to identify measures to be included in the evaluation that are appropriate to the stages of program development at the time the evaluation is being conducted.
- The program theory provides a model against which program activities, assumptions, and goals can be compared through time and that can be used as a basis for identifying changes to the program.
- The program theory can be used as the foundation from which recommended program changes can be made and to estimate the potential effects of those recommendations on the operations or success of the program.
- The program logic model can summarize the key testable hypotheses; each of the outputs and outcomes are possible indicators, and each causal arrow can be tested for its occurrence, its mechanism, and whether there are important external influences on them that also need to be measured. These can form the list of potential measurement elements that are then prioritized and assessed for whether and when they should be measured.
- Permits multiple explanations for how a program is expected to work.
- Enables the incorporation of unintended consequences or new learning about the market and its participants.

Due to the importance of using the program theory and logic model as part of evaluation planning, program theories and logic models that are already developed by the program are critically assessed by the evaluator during the first steps of the evaluation planning process. In this way verification can be made on whether all the necessary elements are there and seem logical given relevant theories and experience in the evaluation field. This is necessary to ensure that the PT/LM can provide an accurate foundation for the evaluation plan.

Program Theories are Especially Important to the Evaluation of Complex Programs

Using a program theory in the evaluation planning process is important for all types of program evaluations. However, the use of a program theory is particularly important for evaluations of complex programs with long-term goals such as information and educational programs, and programs that are trying to change how a market operates, where the program activities and the desired long-term or ultimate outcomes of the program may be many steps removed from one another.

In the case of resource acquisition programs, the causal linkages between an intervention and the desired behavior are usually quite direct and the program theory quite compact. For example, a rebate is offered to reduce the cost of an efficient washing machine to

increase the number of householders purchasing efficient washing machines in order to reduce consumer energy use (*2001 Framework Study*, Chapter 4).

For programs that attempt to change how a market operates (market transformation), the causal chain may be quite lengthy involving numerous interventions in different parts of a market with different market actors. For example, an education program targeting grade schools may depend on a series of events to occur over a considerable period of time before the effects of the program begin to appear. Likewise, a program that attempts to change the way products are manufactured may need to rely more on efforts that change product demand levels beyond what can be achieved by a limited number of rebates offered to one or two segments of the market at a specific point in time. In these cases the program theory may be more elaborate than the theories associated with the resources acquisition program.

The Relationship Between Program Theory, Market Assessment, and Market Theory

In order to develop a useful program theory, an understanding of the market is crucial. For example, the important decision makers for the facilities of major retail chains are often a small group of architects and engineers located in proximity to the retail chain's headquarters who do most of the design work for the chain. These "image designers" may employ local architects and engineers to help steer projects through local zoning and permitting processes but the local architects and engineers have little say in the overall design. If one is attempting to target new construction or major renovations in retail establishments, it is important to understand this relationship. The implication of this is that the local design is being driven by regional or national standards and not by the local architects or engineers who must focus on the permitting process. It may be difficult for locality-based organizations to influence these standards. Regional and national efforts might be more effective in influencing national retailers. In some cases the national firms are aware of the potential for local rebates and may have standard designs that can be used to capture the rebates thus making them free riders for the local program. Local efforts to improve energy efficiency might better be focused on local retailers and the local architects and engineers who serve them. The important point of this example is that without an understanding of the market that one is trying to influence substantial resources can be directed to parts of the market where little or no effects can be expected.

Thus, program theories need to be based on good market theories and good market theories need to be based in strong empirical evidence. In the course of developing a program theory, program designers need to have an understanding of how their target market operates. Where empirical data is lacking to support a market theory, program designers may need to conduct a market characterization study. (See Chapter 10 for a description and content of a market characterization study.)

A program theory is only as good as its empirical underpinnings. If the market theory is absent or wrong, then it is possible that the program theory will be incorrect as well. This

may cause the evaluator to focus on issues that are not relevant and may lead to conclusions about the program that are not helpful.

Evaluation Ethics

Credibility of evaluations and evaluators are absolutely essential for evaluations to fill their role in providing credible findings on the results from the program and for providing recommendations for program refinement and investment decisions. This makes evaluation ethics a critical foundation for evaluations and the field of evaluation. The Framework could not be complete without reinforcing the importance of evaluation ethics in the practice of evaluating energy efficiency programs.

The American Evaluation Association (AEA) has developed and approved a set of guiding principles for evaluators. These well-established guidelines are part of the new California Evaluation Framework and the authors recommend that these principles guide energy program evaluations and the evaluators who conduct these studies. These guidelines are provided in every addition of The American Journal of Evaluation, on the AEA's web site, and were provided in the preface pages of the hard-copy *Proceedings of the 2003 International Energy Program Evaluation Conference*.

These guidelines are provided in their entirety in the Appendix to this chapter. A high level summary of the guiding principles is given below.

- A. Systematic Inquiry: *Evaluators conduct systematic, data-based inquiries about whatever is being evaluated.*
- B. Competence: *Evaluators provide competent performance to stakeholders.*
- C. Integrity/Honesty: *Evaluators ensure the honesty and integrity of the entire evaluation process.*
- D. Respect for People: *Evaluators respect the security, dignity, and self-worth of the respondents, program participants, clients, and other stakeholders with whom they interact.*
- E. Responsibilities for General and Public Welfare: *Evaluators articulate and take into account the diversity of interests and values that may be related to the general and public welfare.*

History of Energy Efficiency Program Evaluation in California

The history of California's energy efficiency programs and related measurement and evaluation (M&E) requirements can be divided into four distinct periods: the pre-Protocol era (1970s-1994), the Protocol era (1994-1997), the Restructuring era (1998-2000), and the current transition period (post-2000). In particular, in the last five years the environment within which energy efficiency programs operate has changed

enormously. The market structure for electricity in California changed, and simultaneously, the California Public Utilities Commission (CPUC) shifted the focus of energy efficiency programs from resource acquisition to market transformation, and a new generation of programs was added to the older ones. However, the energy crisis of 2001 forced regulators to revisit these policy decisions, with a renewed emphasis on resource acquisition. The optimal balance between resource acquisition and market transformation is still in question, and the relative emphasis is expected to change as conditions warrant. The optimal roles of utilities and other market entities also remain to be decided. The CPUC's desire to update the Framework for assessing programs to meet the new realities is the genesis of this report.

1970s-1994: The Pre-Protocol Era

For over twenty years, the CPUC has approved the use of ratepayer funds to promote energy efficiency activities, and authorized the major investor-owned utilities (IOUs) under its jurisdiction to administer a wide variety of energy efficiency programs. By the early 1990s, a wide variety of these programs began to be planned and carried out as part of the biennial resource planning update with the CPUC and the California Energy Commission (CEC). By the early 1990s, energy efficiency programs and other demand-side management (DSM) activities were identified as viable, cost-effective alternatives to supply-side energy generation projects (the “resource acquisition” perspective).

From the late 1980s until 2002, the CPUC allowed the utilities to recover from ratepayers the costs of “shareholder incentive mechanisms.” The terms and conditions under which the utilities were allowed to claim and recover these transfers varied greatly and the earnings claims were verified in separate proceedings, and each mechanism allowed for a very short earnings recovery period.

In June 1990, the CPUC permitted the Division of Ratepayer Advocates (DRA, later ORA) to receive funds to review utilities' energy efficiency programs. During this period, utilities were calculating the results of their programs through engineering estimates (often referred to as “*ex-ante* estimates”). Beginning in 1993, the CPUC required that utilities should rely more on measurement than engineering and required that the evaluations of energy efficiency programs be based on “*ex-post* measurement.” The 1993 CPUC decision also required that measurement and evaluation protocols be established, and formally established CADMAC – the California DSM Measurement Advisory Committee – to develop these protocols. CADMAC prepared semi-annual reports on the progress of the protocols and also hosted informal workshops where participants could freely discuss the technical issues of the protocols and their implications.

1994-1997: The Protocol Period

For the 1994-1997 years, statewide consistency was established for the shareholder incentive mechanisms, more rigorous terms and conditions for the measurement and verification of costs and benefits were established based on the CADMAC-developed

Protocols. These were adopted by the CPUC in D.93-05-063, providing the rules by which impact evaluations were done to determine the energy savings achievements of programs for which shareholder earnings were awarded. Earnings claims were addressed in the consolidated Annual Earnings Assessment Proceeding (AEAP), with a relatively lengthy earnings recovery period of up to 10 years. The protocols required first-year impact evaluations for most programs (emphasizing regression analysis and use of billing data) and persistence studies (fourth and ninth year measure retention studies, and research on relative technical degradation). Since the ultimate earning level depended on the measured persistence of savings, studies under these protocols are still in progress and are scheduled to occur again in 2006.

1998-2000: Energy Efficiency, Electric Industry Restructuring, and The CBEE Period

Beginning in 1998, several significant program design and implementation changes occurred. First, funding for traditional rebates was reduced and replaced by Standard Performance Contract (SPC) programs, where savings and incentives were based on measured performance. Second, funding for “upstream market transformation” interventions was substantially increased. Third, utility performance awards were substantially delinked from cost-effectiveness considerations, and there were reductions in the earnings opportunities for the utilities. Fourth, there were significant increases in expenditures on M&E studies that attempted to quantify market effects and indirect benefits attributable to the expanded upstream market transformation programs.

Beginning in 1998, the CPUC moved to a market transformation goal for programs.²⁵ The CPUC codified a policy that emphasized removing barriers to energy efficiency in the market so that private sector entities would be able to provide energy efficiency services once public monies were no longer available to fund activities.

As a result of this move towards market transformation, an efficiency advisory board, the California Board for Energy Efficiency (CBEE), established a new set of expectations. The CBEE provided advice to the CPUC on the types of programs to fund as well as the type of measurement and evaluation requirements needed for evaluating market transformation programs. During this period, the California Measurement Advisory Council (CALMAC) was formed to address issues related to programs conducted in 1998 and onwards, particularly M&E efforts. M&E efforts focused on verification of the number of energy efficiency measures performed or installed as reported by the IOUs, and the measurement of actual energy savings achieved was de-emphasized.

The CBEE recommended a new cost-effectiveness test, the Public Purpose Test (PPT). The PPT includes elements that were not traditionally included in the TRC calculations, such as spillover savings, non-energy costs/benefits, positive/negative externalities, and reductions in the cost of measures or practices caused by the program. In addition, the PPT is applied at the portfolio level to encourage investment in interventions that may not

²⁵ See *A Scoping Study on Energy Efficiency Market Transformation by California Utility DSM Programs* (* Eto et al. 1996).

produce measurable savings in the early years, but are more cost-effective over time as market effects compound in later years. In April 1999, CPUC Resolution E-3592 approved modifications to the Policy Rules and, in so doing, adopted the PPT as the standard for cost-effectiveness.

One of the last studies requested by CBEE involved a team of evaluation and cost-effectiveness experts who completed a major framework that discussed the justification for, the logic behind, and the techniques needed for evaluating and assessing the cost-effectiveness of market transformation interventions (the *2001 Framework Study*).²⁶ While the results are worthy of serious consideration, they were never incorporated in any set of M&E protocols or guidelines due to the continuing turmoil in DSM planning and regulation in California.

Post-2000: Transition Period and the Energy Crisis

The current transition period has been a period of great uncertainty, sparked by the Energy Crisis of 2001.

The Energy Crisis

By almost any measure, the events surrounding the electricity situation in California in the 2000/2001 time period were simply extraordinary.²⁷ Between the summer of 2000 and the early winter months of 2001, the California Independent System Operator declared over seventy days of system emergencies, and rolling blackouts were actually initiated on several occasions. In January and February 2001, the CEC projected electricity supply and demand for the summer of 2001 under various temperature scenarios, and analyses suggested that the State could face a potential shortfall of five thousand megawatts during the months of June through September.

In reaction to this, California added significant new funding for energy efficiency.^{28, 29} In all, policy makers ordered more than a 250% increase in spending over the 2000 level. Most of this funding focused on resource acquisition activities to address the near-term crisis, rather than market transformation.

Changing Rules

In November 2001, the CPUC changed the rules for energy efficiency programs to allow non-utilities to compete with utilities for energy efficiency funding in Decision 01-011-066. The Commission eventually selected approximately 40 third parties to administer approximately \$100 million in funds for local 2002 and 2003 energy efficiency programs.

²⁶ *A Framework for Planning and Assessing Publicly Funded Energy Efficiency* (* Sebold et al. 2001).

²⁷ *Examining California's Energy Efficiency Policy Response to the 2000/2001 Electricity Crisis: Practical Lessons Learned Regarding Policies, Administration, and Implementation.* (Kushler and Vine 2003).

²⁸ *Evaluation: Methods for Studying Programs and Policies* (* Weiss 1998).

²⁹ "Using Energy Efficiency to Help Address Electric System Reliability: An Initial Examination of 2001 Experience." (Kushler et al. 2003)

For 2002, the CPUC adopted a new Energy Efficiency Policy Manual (contained in CPUC Decision 01-011-066),³⁰ and program evaluation and numerous other program activities are governed by this set of protocols. The manual contains the requirements for cost-effectiveness analysis, and requires measurement and verification using the U.S. Department of Energy's International Measurement and Verification Protocol.³¹

For Program Years 2003 and 2004, the penalty “stick” replaced the incentive “carrot.” There are no shareholder earnings for either utility- or non-utility-administered energy efficiency programs. Instead, a holdback mechanism has been introduced for the non-utility-administered programs, and utilities are subject to a penalty process for sub par program performance.

Resource Procurement and Portfolio Management

In another major policy shift, the CPUC required (in October 2002, Decision 02-10-062) that California's three electric IOUs take responsibility for resource procurement to maintain the reliability of California's electric grid. To meet this CPUC mandate, the utilities filed portfolio management plans in April 2003, which included \$700 million in energy efficiency spending over the next five years in addition to current commitments through the Public Goods Charge funds. With such a mix of implementers and policy objectives, new evaluation protocols are simply one of the many pieces that must come together to create a consistent and reliable program environment.

Policy Use of Evaluation Results

Policy Based on Evaluation Results

Good policy is based upon reliable, accurate information. It is this principle that makes quality evaluations imperative. Its importance can be seen in the current movement towards greater “evidence-based decision-making” in policy arenas^{32, 33} and health care.³⁴ Evidence-based decision-making requires a focused question that is then addressed by assessing and applying the evaluation results to answering that question through the art of critical thinking. It is important for this process to include information concerning uncertainty and potential bias. (See Chapter 12: Uncertainty, for a discussion of the Framework's guidelines for these issues.) With a systematic approach using these methods, policy can be well-grounded, well-supported, and better allow consistent and

³⁰ (CPUC 2001).

³¹ (International Performance Measurement and Verification Protocol 2001).

³² See (Public Policy Forum 2003) for information on UK Government movement on evidence-based decision-making.

³³ “Evaluation, Policy Learning and Evidence-Based Policy Making.” (Sanderson 2002).

³⁴ A large effort of the last 5-10 years has developed here for clinical applications. Evidence-based decision-making gained prominence in Canadian health care with the Prime Minister's National Forum on Health in 1997, where it was defined as “the systematic application of the best available evidence to the evaluation of options and to decision-making in clinical, management, and policy settings” from *Health Services Research and Evidence-Based Decision-Making* (Canadian Health Services Research Foundation 2000).

incremental policy that is refined as research suggests, increasing the likelihood that the acceptance and adjustments to policy changes can occur more smoothly.

Interpreting and Adjusting Results for Summative versus Formative Purposes

It is important that users of evaluation results interpret study results correctly. As discussed earlier in this chapter, a summative evaluation answers the question of what occurred within a program over a particular time period. This information is appropriate for use in understanding the program's accomplishments and its contribution to the overall portfolio. Further, information from a formative evaluation is needed to decide what that program's future results might look like, indicating whether it should be continued or not. Similarly, summative results only provide partial information on which to base a decision about offering a program in a different location. In this latter case, formative evaluation information is needed to understand why and under what conditions the results can be obtained in order to help determine if the program should be replicated at another locations and if the results from the new program would be similar or different.

An example of this concept that is used frequently within this Framework is a program that is in its early implementation stage that has significant start-up costs. If during the first year a benefit-cost analyses is used to decide whether to continue the program a significant risk occurs that the program might be highly cost-effective over the long-term, but cancelled because of the results of the first year assessment. This means the initial savings and benefit-cost analysis may be one of several important elements to report as part of the summative evaluation. However, there should also be a formative assessment concerning the probable long-term benefit-cost that may provide a different picture of the program to policy makers.

A less obvious example is when the effects of a significant change in the economy also affects the impact evaluation results. For example, in cases when a recession coincides with post-retrofit periods, a billing analysis may overstate savings. In this case correction techniques should be applied in the billing analysis to capture this effect and correct the energy savings estimates. Engineering methods that incorporate fieldwork may recognize this condition and give the evaluation team a notice of the current building condition. At the same time, both methods need to provide an estimate of what savings would be expected under more steady-state economic conditions. A similar but opposite challenge might occur during periods of fast economic growth. In these cases, crowding of employees might occur prior to physical expansion (by either new construction or renting new space). This could mean that savings are actually greater than what would show up in a billing analysis. Billing analysis, that doesn't correct for this condition, will understate savings (as post-retrofit usage is greater). Engineering estimates of savings can overstate long-term savings if the current operating conditions are assumed to be occurring over the long-term, when instead they may actually only be a temporary condition. Each of these scenarios points to the need for careful evaluation with accurate analysis and reporting that reflects changes going out beyond the energy technology changes.

Summative results also need to be assessed in conjunction with process evaluation results in order to understand the program's potential under differing program and operational conditions to support decisions associated with program continuation or funding.

Making the Framework a Living Document

This document provides a great many references to help guide policy makers, energy suppliers, evaluation planners, evaluators, and other stakeholders understand the environment in which evaluations can be conducted and the scope of these efforts. However, as evaluation efforts are undertaken evaluation professionals will continue to find ways to improve upon the current state-of-the art. As additional research is completed and improvements are discovered for specific evaluation practices, a systematic process needs to be in place that will allow the findings from this research to be used to modify and refine the Framework, either as an actual document modification allowing for the Framework to be updated, or through a referral sending users to the potentially modifying documents.

Similarly, as elements of this Framework are used in the field, lessons will be learned that can lead to additional recommendations for refining this Framework. Likewise, a systematic process needs to be put in place that encourages and allows this evolution to occur.

Finally, the evaluation field is not static. The highest quality Framework needs to be able to change and grow in response to developments within the energy program evaluation field as well as the larger field of evaluation. A means for this to occur needs to be established.

Appendix to Chapter 4

An Example of a Program Theory

To illustrate a program theory, we examine a program with a goal to reduce energy consumption in multifamily buildings. The program targets owners of multifamily buildings who are making equipment replacement decisions typically for HVAC systems, including large air conditioning systems and central or distributed heating systems. The idea is to obtain immediate savings from owners who are currently making equipment replacement decisions by encouraging more efficient equipment installations, while encouraging these owners, their contractors, and their distributors to adopt more energy efficient decision practices and equipment in the longer term. In this example the program is operated via a program vendor who works with equipment installation contractors to identify facility owners who are or can be encouraged to be interested in an energy efficient assessment of their facility linked with rebates for efficient equipment replacements or upgrades that are recommended as a result of the assessment.

A portion of the program theory for this program might look something like the following:

Multi-family market research conducted by the program vendor indicates that facility owners are interested in increasing or maintaining the value of their property, reducing the hassles associated with facility maintenance, reducing the operational costs associated with the facility, being seen in the market as a provider of high-quality rental units, and making the facility more attractive to potential buyers. This market condition results in a market opportunity for a program to offer services that makes it economically attractive for owners to achieve their goals while saving energy at the same time. The program vendor hypothesizes that if they work through equipment installation contractors who have pre-existing relationships in the multi-family market they can use these relationships to convince owners to make the changes needed to achieve the program goals.

The market theory indicates that there are two multi-family market sub-segments that can be successfully reached by the program that will produce energy savings. These include:

1. Facility owners who are not currently considering an upgrade but who can be convinced to retrofit their facilities with more efficient equipment, and
2. Facility owners who are considering an upgrade to install low-cost/low-efficiency equipment, but who can be convinced to install more efficient equipment than they were originally considering.

The program vendor hypothesizes that through these two types of market actors the program can deliver a specific level of energy and demand savings to California's energy markets. The program vendor indicates that they can capture the energy savings through the following program activities based upon the program assumptions concerning the

market and the consequences (outputs and the causal chain of outcomes from these outputs).

Program Activities

- Initial program launch will have the vendor working with a small set of innovative customers already known by the vendor. Vendor will establish relationship with these customers as the innovative launch customers.
- The program vendor will make personal contact with installation contractors and describe in detail the benefits and operations of the facility assessment and reward program.
- The program vendor will design, develop, and provide to the contractors program marketing and outreach materials.
- When a customer seeking an inefficient replacement system approaches an installation contractor or when the contractor judges that a customer or potential customer is a good candidate for the program, the installation contractor will contact the customer and recommend the facility assessment and reward program.
- The installation contractor will then instruct the program vendor to conduct the assessment, recommending equipment and/or installation practices that will save energy and accomplish the customer's participation objectives. The program vendor makes a presentation to the owner and the installation contractor offering a reward that represents the incremental cost between what would have been the standard installation and the efficiency upgrade.
- The program vendor assists the installation contractor as necessary, concerning where efficiency equipment can be obtained and the methods for its proper installation.
- The owner and the installation contractor confer and the owner decides to proceed with the upgrade.
- The upgrade is installed.

Program Assumptions

- The program vendor can organize and implement the program and make contacts with the target customers.
- Initial participants will be a small set of innovative customers known by the vendors, and these relationships with these customers will help launch the program.
- The program marketing and outreach materials (developed by the vendor) will be accepted by the contractors and help gain acceptance by their customers in such a way as to increase program enrollments and equipment installations.

- The vendor's activities with these contacts and the program marketing materials will be enough to motivate contractors to join the program and market to their customers and potential customers.
- Contractor enrollments and participation will occur at the rate assumed from both targeted market segments such that the program achieves its energy-related goals.
- The program is attractive to installation contractors because it gives them a new tool to use to gain additional business, added profits, and additional customers than they would have without the program.
- The program is attractive to customers because it provides a free assessment service and cash rewards to customers in two market segments that enables them to increase the value and marketability of their property while reducing operational costs.
- The market theory describes several market barriers for contractor enrollments and customer participation, but identifies several market and technology advantages associated with the program that the vendor thinks could overcome those barriers.
- The installations are made in a high quality manner and assumed energy and demand savings are thus achieved.

The goal of this program is defined in the program theory to reduce the energy consumption of 100 multi-family facilities having a total of more than 1,500 living units in the South Bay area of California by "X" kWh a year with a corresponding drop in coincident demand by "X" kW during the first year of program operation, by "X" kWh a year with a corresponding drop in coincident demand by "X" kW during the second year of program operation, and by "X" kWh a year with a corresponding drop in coincident demand by "X" kW during the third year of program operation. The program theory indicates that eighty percent of the projected savings will be produced over a fifteen-year period consistent with the expected life of the installed technology.

What these assumptions and activity statements represent is a series of hypotheses about how the program vendor thinks the program will work in the market place. A logic model can be created that represents how the program activities are hypothesized to lead to program outputs. Then there are assumptions of how the program outputs will lead to short-term outcomes. The first activities statement hypothesizes that personal relationships between the vendor and a small number of innovative contractors is important to the program launch. Another hypothesis is that the vendor's work with contractors and the program marketing materials will cause contractors to participate with the program and be able to "sell" their customers on participating in the program. These and all of the other assumptions and program logic included within the above become the evaluation's testable hypotheses.

An Example of a Logic Model

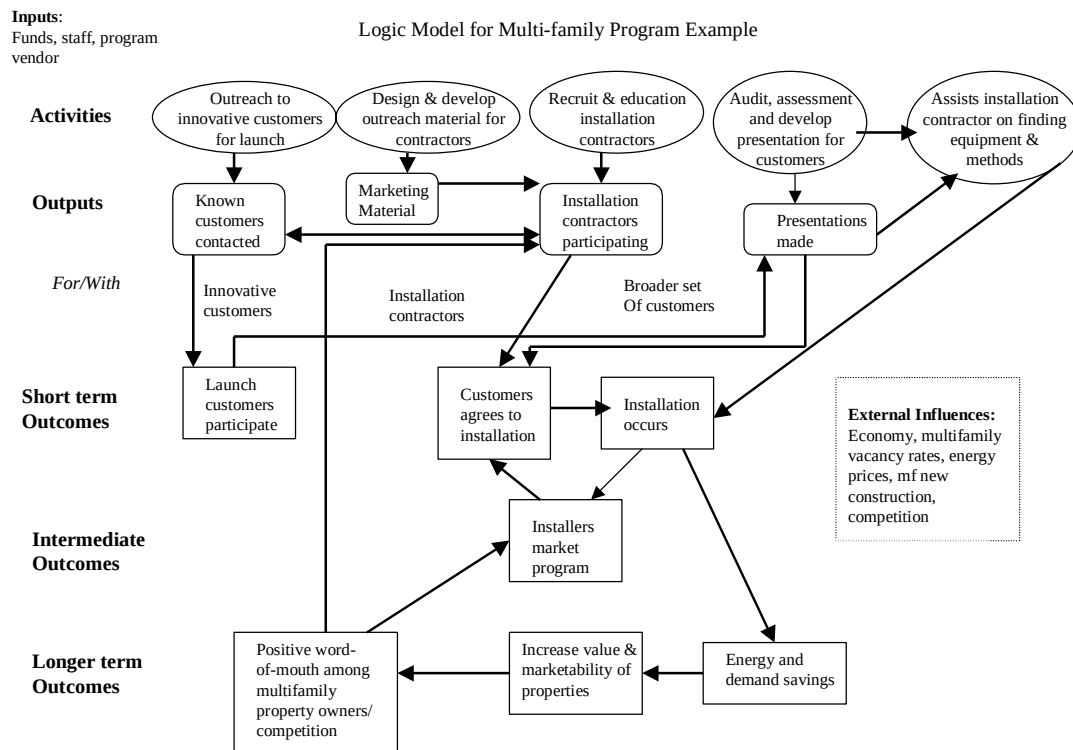


Figure 4.2: Logic Model for Multi-family Program Example

Complete Evaluation Ethics Document from the American Evaluation Association

These guidelines are copied here in their entirety as printed in The American Journal of Evaluation.

American Evaluation Association: Guiding Principles for Evaluators

The American Evaluation Association (AEA), the professional association for evaluators, works to ensure ethical work in the evaluations of programs, products, personnel, and policy. To proactively guide the work of professionals in everyday practice and to inform evaluation clients and the general public of expectations for ethical behavior, the Association developed the following guiding principles:

- A. Systematic Inquiry: *Evaluators conduct systematic, data-based inquiries about whatever is being evaluated.*
1. Evaluators should adhere to the highest appropriate technical standards in conducting their work, whether that work is quantitative or qualitative in nature, so as to increase the accuracy and credibility of the evaluative information they produce.
 2. Evaluators should explore with the client the shortcomings and strengths both of the various evaluation questions it might be productive to ask, and the various approaches that might be used for answering those questions.
 3. When presenting their work, evaluators should communicate their methods and approaches accurately and in sufficient detail to allow others to understand, interpret, and critique their work. They should make clear the limitations of an evaluation and its results. Evaluators should discuss in a contextually appropriate way those values, assumptions, theories, methods, results, and analyses that significantly affect the interpretation of the evaluative findings. These statements apply to all aspects of the evaluation, from its initial conceptualization to the eventual use of findings.
- B. Competence: *Evaluators provide competent performance to stakeholders.*
1. Evaluators should possess (or, here and elsewhere as appropriate, ensure that the evaluation team possesses) the education, abilities, skills, and experience appropriate to undertake the tasks proposed in the evaluation.
 2. Evaluators should practice within the limits of their professional training and competence, and should decline to conduct evaluations that fall substantially outside those limits. When declining the commission or request is not feasible or appropriate, evaluators should make clear any significant limitations on the evaluation that might result. Evaluators should make every effort to gain the competence directly or through the assistance of others who possess the required expertise.
 3. Evaluators should continually seek to maintain and improve their competencies, in order to provide the highest level of performance in their evaluations. This continuing professional development might include formal coursework and workshops, self-study, evaluations of one's own practice, and working with other evaluators to learn from their skills and expertise.

C. Integrity/Honesty: *Evaluators ensure the honesty and integrity of the entire evaluation process.*

1. Evaluators should negotiate honestly with clients and relevant stakeholders concerning the costs, tasks to be undertaken, limitations of methodology, scope of results likely to be obtained, and uses of data resulting from a specific evaluation. It is primarily the evaluator's responsibility to initiate discussion and clarification of these matters, not the client's.
2. Evaluators should record all changes made in the originally negotiated project plans, and the reasons why the changes were made. If those changes would significantly affect the scope and likely results of the evaluation, the evaluator should inform the client and other important stakeholders in a timely fashion (barring good reason to the contrary, before proceeding with further work) of the changes and their likely impact.
3. Evaluators should seek to determine, and where appropriate be explicit about, their own, their clients', and other stakeholders' interests concerning the conduct and outcomes of an evaluation (including financial, political, and career interests).
4. Evaluators should disclose any roles or relationships they have concerning whatever is being evaluated that might pose a significant conflict of interest with their role as an evaluator. Any such conflict should be mentioned in reports of the evaluation results.
5. Evaluators should not misrepresent their procedures, data, or findings. Within reasonable limits, they should attempt to prevent or correct any substantial misuses of their work by others.
6. If evaluators determine that certain procedures or activities seem likely to produce misleading evaluative information or conclusions, they have the responsibility to communicate their concerns, and the reasons for them, to the client (the one who funds or requests the evaluation). If discussions with the client do not resolve these concerns, so that a misleading evaluation is then implemented, the evaluator may legitimately decline to conduct the evaluation if that is feasible and appropriate. If not, the evaluator should consult colleagues or relevant stakeholders about other proper ways to proceed (options might include, but are not limited to, discussions at a higher level, a dissenting cover letter or appendix, or refusal to sign the final document).
7. Barring compelling reason to the contrary, evaluators should disclose all sources of financial support for an evaluation, and the source of the request for the evaluation.

D. Respect for People: *Evaluators respect the security, dignity, and self-worth of the respondents, program participants, clients, and other stakeholders with whom they interact.*

1. Where applicable, evaluators must abide by current professional ethics and standards regarding risks, harms, and burdens that might be engendered to those participating in the evaluation; regarding informed consent for participation in evaluation; and regarding informing participants about the scope and limits of confidentiality. Examples of such standards include federal regulations about protection of human subjects, or the ethical principles of such associations as the American Anthropological Association, the American Educational Research

Association, or the American Psychological Association. Although this principle is not intended to extend the applicability of such ethics and standards beyond their current scope, evaluators should abide by them where it is feasible and desirable to do so.

2. Because justified negative or critical conclusions from an evaluation must be explicitly stated, evaluations sometimes produce results that harm client or stakeholder interests. Under this circumstance, evaluators should seek to maximize the benefits and reduce any unnecessary harm that might occur, provided this will not compromise the integrity of the evaluation findings. Evaluators should carefully judge when the benefits from doing the evaluation or in performing certain evaluation procedures should be foregone because of the risks or harms. Where possible, these issues should be anticipated during the negotiation of the evaluation.
3. Knowing that evaluations often will negatively affect the interests of some stakeholders, evaluators should conduct the evaluation and communicate its results in a way that clearly respects the stakeholders' dignity and self-worth.
4. Where feasible, evaluators should attempt to foster the social equity of the evaluation, so that those who give to the evaluation can receive some benefits in return. For example, evaluators should seek to ensure that those who bear the burdens of contributing data and incurring any risks are doing so willingly, and that they have full knowledge of, and maximum feasible opportunity to obtain any benefits that may be produced from the evaluation. When it would not endanger the integrity of the evaluation, respondents or program participants should be informed if and how they can receive services to which they are otherwise entitled without participating in the evaluation.
5. Evaluators have the responsibility to identify and respect differences among participants, such as differences in their culture, religion, gender, disability, age, sexual orientation, and ethnicity, and to be mindful of potential implications of these differences when planning, conducting, analyzing, and reporting their evaluations.

E. Responsibilities for General and Public Welfare: *Evaluators articulate and take into account the diversity of interests and values that may be related to the general and public welfare.*

1. When planning and reporting evaluations, evaluators should consider including important perspectives and interests of the full range of stakeholders in the object being evaluated. Evaluators should carefully consider the justification when omitting important value perspectives or the views of important groups.
2. Evaluators should consider not only the immediate operations and outcomes of whatever is being evaluated, but also the broad assumptions, implications, and potential side effects of it.
3. Freedom of information is essential in a democracy. Hence, barring compelling reason to the contrary, evaluators should allow all relevant stakeholders to have access to evaluative information, and should actively disseminate that information to stakeholders if resources allow. If different evaluation results are communicated in forms that are tailored to the interests of different stakeholders, those communications should ensure that each stakeholder group is aware of the

- existence of the other communications. Communications that are tailored to a given stakeholder should always include all important results that may bear on interests of that stakeholder. In all cases, evaluators should strive to present results as clearly and simply as accuracy allows so that clients and other stakeholders can easily understand the evaluation process and results.
4. Evaluators should maintain a balance between client needs and other needs. Evaluators necessarily have a special relationship with the client who funds or requests the evaluation. By virtue of that relationship, evaluators must strive to meet legitimate client needs whenever it is feasible and appropriate to do so. However, that relationship can also place evaluators in difficult dilemmas when client interests conflict with other interests, or when client interests conflict with the obligation of evaluators for systematic inquiry, competence, integrity, and respect for people. In these cases, evaluators should explicitly identify and discuss the conflicts with the client and relevant stakeholders, resolve them when possible, determine whether continued work on the evaluation is advisable if the conflicts cannot be resolved, and make clear any significant limitations on the evaluation that might result if the conflict is not resolved.
 5. Evaluators have obligations that encompass the public interest and good. These obligations are especially important when evaluators are supported by publicly generated funds; but clear threats to the public good should never be ignored in any evaluation. Because the public interest and good are rarely the same as the interests of any particular group (including those of the client or funding agency), evaluators will usually have to go beyond an analysis of particular stakeholder interests when considering the welfare of society as a whole.

Chapter 5: Umbrella Framework for Evaluation Planning, Funding, and Research Prioritization Decisions

Introduction

This chapter presents and discusses the evaluation planning cycle, the program evaluation goals placed within that cycle, and issues relating to the planning and review of the evaluation plan. The chapter then presents an example of an evaluation approach integrated within a program planning and implementation cycle. These sections allow the reader to understand the structure of the program cycle in which the evaluation process is placed.

The chapter then focuses on the evaluation planning and development process, including a review of the key issues associated with the evaluation planning and review process, the completion of the evaluation plan, and issues relating to the plan review and approval process. The chapter then displays a high-level diagram of the Framework, showing the relationships of the different parts of the Framework. The chapter then discusses evaluation priorities, budgets, and opportunities for evaluation consolidation.

Finally, the authors have included an appendix to the chapter providing a more detailed example of an integrated “systems approach” program implementation and evaluation cycle.

The primary goal of this document is to set the stage for providing guidance on what types of evaluations are appropriate in California and to provide guidance to those responsible for preparing and approving evaluation plans, for conducting the evaluations and for reporting the results. The Framework is an evaluation guidance document and is not a prescriptive protocol for conducting evaluations in California. However, the California Public Utilities Commission (CPUC) or another designated organization charged with the responsibilities of developing new evaluation protocols may wish to use this Framework as one of the guiding documents for establishing new evaluation protocols for California’s energy efficiency programs.

As with other chapters in this document, the umbrella chapter is written for a wide range of stakeholders within the California energy efficiency program arena. However, because of the topics covered in this chapter and the impact of this chapter on the development and implementation of a statewide evaluation framework, a wide range of affected stakeholders should have a working knowledge of the structure and issues described in this chapter, including:

- Regulatory staff & policy makers responsible for evaluation;
- Evaluation oversight managers & reviewers;
- Program administrators and program implementers;
- Cost-effectiveness and avoided cost analysts;

- Evaluation designers and managers;
- Evaluators for all types of evaluation—impact, metering and monitoring, process, information and education programs, market transformation and market effects, and non-energy effects;
- Statisticians & research data managers involved in evaluations; and
- Portfolio managers.

The Evaluation Planning Cycle

The evaluation processes and roadmaps associated with this Framework are placed within a cyclic evaluation-planning process consistent with the program funding and contracting cycles. This Framework takes no position on whether those cycles should be one, two, three, or more year cycles, but does suggest that the evaluation planning cycle be consistent with the program approval and funding cycles. This allows the program evaluation planning process to structure the evaluation plans so that the evaluations are planned, implemented, and completed within a program cycle. Later in this chapter we discuss one approach for coordinating the program implementation and evaluation cycles in order to illustrate the need to coordinate the two processes over the same period of time.

For two-year cycles the program evaluation plans would cover all the evaluation efforts that are planned over the two years of the program contract period. Likewise, a three-year program cycle would mean the evaluation planning process cover a three-year period. The evaluation plan would then specify the types of evaluation efforts that are scheduled throughout the program period. The evaluation planning process would take into consideration the need for evaluation results to document the operations and effects of the program, provide feedback for ongoing program improvement, provide information to support energy efficiency portfolio assessments, and to help support the planning efforts for future program cycles. For impact evaluations that examine technology-specific impacts (rather than program impacts), the information can also be used to update the DEER. The timelines and processes for these efforts can be part of the evaluation protocol formation process that may be developed under the direction of, and adopted by, the CPUC.

The Framework process does suggest that program administrators plan their evaluation efforts over the duration of the program period. This may take the form of an annual evaluation plan for each year of the program cycle, or a multi-year plan with detailed evaluation efforts presented for each year of the program.

The overall plan for the evaluations of a program should describe the types of studies that are proposed for the program, the justification for why the different types of evaluations are being proposed and why others are not, the timeframe for these studies, a presentation of the CPUC's evaluation goals that each study addresses, and a description of how each goal will be addressed.

These current CPUC Policy Manual³⁵ evaluation goals are to:

1. Measure energy and peak savings;
2. Measure cost-effectiveness;
3. Provide upfront market assessment and baseline analysis;
4. Provide ongoing feedback and guidance to the program administrator;
5. Measure indicators of effectiveness and testing program theory and approach;
6. Assess the overall levels of performance and success;
7. Inform decisions regarding compensation and final payments; and
8. Help assess the continuing need for the program.

These specific goals may be modified from time to time to support changes in public policy or to fine-tune the goals so that they provide more defined instructions to support the evaluation planning process. However, detailed evaluation plans for individual programs should include the following information, as appropriate for the type of evaluation being conducted:

1. The type of evaluations being conducted, what CPUC evaluation goals it will meet, and the logic and support for the proposed set of evaluation studies and why these have higher priority over other types of evaluation (whenever the evaluation budget limits conducting all of the evaluations that would be desired);
2. A short description of the program(s) being evaluated;
3. A presentation of the program theory;
4. A short description of the expected program effects;
5. A list of the technologies offered by the program with an indication of which technologies will be addressed in the evaluation(s);
6. A presentation of the researchable issues to be addressed in the evaluation(s) and reported in the evaluation reports;
7. A presentation of the overall scope of the evaluation reported by type of evaluation (impact, process, market effects, etc.);
8. A detailed presentation of the evaluation activities to be undertaken, by program evaluation type (impact, process, market effects, etc.) including the monitoring, metering and verification efforts to be employed;
9. A presentation of the sampling approach and sample selection methods for each evaluation activity that includes sampling efforts. Sample sizes should be reported at the technology level and summarized to the program level;
10. A description of how the control group, comparison group, or non-participant information will be used in the evaluation(s) and in the analysis;

³⁵ *Energy Efficiency Policy Manual, Version 2.* (* CPUC 2003), page 26.

11. A description of how spillover and free rider effects will be addressed in the evaluation activities and in the data analysis;
12. A brief presentation and discussion of the threats to validity, potential biases, methods used to minimize bias, and the level of uncertainty associated with the sample selection methods and the evaluation approaches;
13. A description of the data handling and data analysis approach to be used to address the researchable issues;
14. A statement of whether the evaluation can be used to update and/or validate specific values, including measure costs, measure life, and measure impacts, within the DEER. This statement should identify the specific DEER technologies and technology characterizations that will be addressed. If the evaluation cannot be used to update the DEER information, the reasons should be cited;
15. An activities timeline with project deliverable dates; and
16. A presentation of the evaluation costs for each type of evaluation (process, impact, metering and verification, market effects, etc.) and the overall costs for all evaluation efforts proposed.

The Framework does not include a description of the process associated with the submission, review, and approval of program-specific evaluation plans. However, the Framework is established with an understanding that there will be a formal process for reviewing and approving evaluation plans to make sure that they provide the level of rigor required to support a portfolio assessment and planning process, and to reliably understand the energy resources available through different types of energy efficiency programs, technologies, and behaviors. The formal review and approval process for evaluation plans will be developed and described by the CPUC.

Components of the Evaluation Plan

The program evaluation plan is a formal document that clearly presents the evaluation efforts and details the activities to be undertaken during the evaluation. The evaluation plan is a stand-alone decision document, meaning it needs to contain the information needed to allow the evaluation plan reviewer to make a decision regarding if the evaluation plan should be approved and implemented or modified. There are several components of the evaluation plan that are needed for the review process, these include the following components:

- Cover page containing the program name, program administrator, evaluation contractor, date of evaluation plan, and the program tracking number (if any);
- Table of Contents;
- Summary overview of the program and the evaluation effort;
- Presentation of the evaluation goals and researchable issues addressed in the evaluation;
- Brief description of the program(s) being evaluated;

- Description of how the evaluation addresses each of the CPUC evaluation objectives (currently as in the CPUC Energy Efficiency Policy Manual³⁶ Chapter 6 EM&V Objectives),
 - Where an evaluation plan prioritizes evaluation needs within a constrained evaluation budget such that not all evaluation objectives are met with completed evaluation studies, then the evaluation plan should provide evidence to support its prioritization to include, as appropriate:
 - comparative program evaluation findings,
 - the history of evaluation findings for the type of program, and
 - a discussion of program maturation and stability/changes;
- Brief presentation of the program theory and underlying assumptions;
- Task description of the evaluation efforts;
- Description of the analysis activities and approaches;
- Description of the M&V efforts (impact evaluations only) including:
 - Reference to IPMVP option, if used, and
 - Description of deviation from IPMVP, if any;
- Description of the sampling rationale, methods, and sample sizes;
- Discussion of the uncertainty and bias issues associated with the evaluation approach and methods incorporated to minimize bias and uncertainty;
- Evaluation activities timeline;
- Program budget and the evaluation budget; and
- Program administrator and evaluation manager contact information.

The evaluation plan should be written in a style and with enough detail that it can be clearly understood by program administrators, policy makers, and evaluation professionals.

Reviewing and Approving the Evaluation Plan

It is important for evaluations to focus on the key research goals for the types of programs being implemented and to focus research on the types of information needed at various times along the program path. It is the evaluation plan that details the researchable issues that are addressed in the evaluation effort and also presents the methods and approaches that are expected to provide the needed information. Policy makers and regulators need to be assured that the evaluations conducted will deliver the type and quality of information needed. Evaluations that are under-designed can waste valuable resources by not being able to reliably provide the information needed or by providing information that is not needed. Likewise, evaluations can also be over-designed to address issues that are not priority issues or employ research methods that can be substituted with less costly approaches. In addition, there is a need to coordinate the overall evaluation approach so that limited evaluation resources can be focused on the issues of importance across the collection of evaluation activities. As a result, there is a need to have a process for reviewing and approving evaluation plans. While the

³⁶ (* CPUC 2003).

Framework does not address the organization and structure of how that review and approval process should be provided, it is important to have a process that guides the evaluation research being conducted to meet the overall policy and portfolio assessment needs as well as the program-specific evaluation needs. The evaluation plan review process could be developed and directed to address the following issues in the review process.

- Does the evaluation address the key policy, regulatory, and oversight needs for evaluation information? Are there key issues missing? How will the missing issues be addressed? Can the approach be strengthened to better address these needs?
- Does the evaluation plan represent a reasonable approach to addressing the information needs and the researchable issues?
- Is the evaluation capable of providing reliable conclusions on energy and other impacts? Are there key threats to the validity of the conclusions? Are they being minimized given budget constraints and study tradeoffs? Will they be documented and analyzed?
- Are the researchable issues clearly presented and described?
- Are there missing opportunities associated with the evaluation approach that should be added or considered?
- Does the cost of the study match the methods and approaches planned?
- Does the evaluation plan make sense as an independent evaluation study, given the other evaluation efforts planned for the same time period for other programs and for the program portfolio (i.e., or does a different complementary or consolidated study provide a better use of evaluation funds)?

Evaluations to Facilitate an Integrated Planning Process

The timing of evaluation studies can be viewed in the context of the overall strategic cycle associated with program planning, implementation, and evaluation.³⁷ In a strategic framework approach, evaluation results are used to make informed decisions on program improvements and future program designs and offerings. The program implementation cycle is one in which programs are designed, then implemented, and then evaluated. Following the results of the evaluation, programs are re-examined for design changes and then modified so that those design changes result in improved program implementation efforts. This cycle provides for a continuing process of program improvement, so that the programs match available market opportunities and continually improve their cost-effectiveness over time.

The timing of these activities is influenced by several, often competing considerations. These considerations include:

³⁷ “DSM Resource Planning the Next Generation: Building the Foundation Through Evaluation.” (Rufo 1993).

- Input to policy considerations and decisions;
- Early feedback to program implementers;
- Program lifecycle stage;
- Evaluation data time lags;
- Portfolio planning requirements;
- Evaluation planning requirements and regulatory oversight;
- Program design/solicitation, selection, review, and implementation preparation;
- The value of dispersed timing for evaluations;
- Contract requirements for “pay for performance” programs;
- Market inertia;
- Timing needs for retention, measure life, and technical degradation analyses; and
- Regulatory oversight and review.

Each of these aspects and their relationship to the evaluation process is briefly discussed in the following paragraphs.

Input to Policy Considerations and Decisions

One of the most important goals of the evaluation effort is to provide policy makers and portfolio decision makers with the information they need to identify and select which programs and program portfolios to offer as an alternative energy resource to the people of California. Some policy makers argue that this is the most important reason why evaluations are funded. Policy makers and portfolio managers need reliable evaluation results to support the policy decisions and their associated timelines. Likewise, portfolio managers need information when they are designing the mix of program services to provide and selecting the target markets for these services. Program administrators and evaluators should consider these needs when designing and conducting evaluation efforts.

Early Feedback to Program Implementers

Program managers benefit from evaluation results as early in the implementation process as practical in order to make timely program design and delivery adjustments to improve the cost-effectiveness of their programs. Process evaluation and Measurement and Verification (M&V) activities conducted early in the program cycle can identify potential problems with program operations and operational procedures as well as hardware selection and installation issues, allowing program implementers to correct these problems early on.

At the same time, early scheduling of process evaluation and the M&V efforts may complicate the sample design process. Participant samples for on-site or phone surveys may need to be drawn in stages, before the full participant population is established. If problems are identified and corrected, then follow-up surveys may be required to verify that program changes are effective.

Program Lifecycle Stage

Programs need sufficient time to start up and establish efficient operation. Implementation contractors, subcontractors, and trade allies need time to “get up on the learning curve” to the point where the program is operating at maximum efficiency. The market may also require stability in program presence and operation before steady-state operation can be achieved. This may be especially important with new programs, but it is also important to allow for a period of operational presence in the market with continuing programs that have implemented significant program design or delivery changes. However, as indicated in the Process Evaluation Chapter, it is also important to have the evaluation team start their work during the early phases of the program design and delivery period to help the program be structured in a way that collects needed evaluation information. Early evaluator involvement can also provide, where appropriate, early reviews of program designs and intended operational systems and procedures and allow for early implementation feedback so that needed program changes can be made as early as possible.

Established and continuing programs may not require start-up periods at the beginning of each program cycle. For these programs, unless there are major changes in the program design and operational procedures, the timing of the process and M&V studies may be less critical. Indeed, for continuing programs without major changes in program design and operational procedures, evaluation during each program cycle may not be an efficient use of resources.

Evaluation Data Time Lags

For some programs, especially small programs or new programs, there is a natural time lag between the completion of a program and the completion of the evaluation study. When a program is small enough or new enough that the evaluation needs to obtain information from participants entering the program during the end of the program cycle, evaluation results may be provided at a later time compared with other evaluations in which the program has enough participants to complete the evaluation efforts earlier. In these cases, the evaluation results may not be available until during or after the planning process for a future version of the program.

Portfolio Planning Requirements

Planning a portfolio of programs requires an understanding of the overall goals of the portfolio, the technical and market potential of various measures and delivery mechanisms to deliver energy and demand savings, the cost-effectiveness of various delivery mechanisms, and the risk and uncertainty associated with various technologies and program approaches. Evaluations can provide much information to support this process, provided the information is delivered in a timely manner and in a format compatible with the portfolio analysis process. Market saturation studies, which inform technical and market potentials, should be conducted on a schedule that provides timely information to the planning process.

Portfolio planning can be considered an ongoing process. Portfolio objectives can be updated with the best available information during each planning cycle. Portfolio objectives will likely change as policies change, markets change, and the saturation of energy efficient technologies increases. The portfolio plan can be viewed as a “living document,” subject to continuous refinement and revision.

Evaluation Planning Requirements and Regulatory Oversight

Evaluation studies need to be planned from a portfolio perspective, considering the portfolio objectives, risks, and evaluation data needs. Overall evaluation goals and objectives, evaluation priorities, and program design issues should be addressed prior to planning individual program evaluation studies. Issues such as evaluation resource allocation, evaluation study consolidation, the need to evaluate a particular program during a given program cycle, and evaluation scheduling should be considered prior to planning individual evaluation studies. After this, time needs to be scheduled for the processes of regulatory review and approval, selection of evaluation contractors, and development of detailed evaluation plans. The schedule of these activities will likely be defined by the prevailing administrative structure.

Program Design/Solicitation, Selection/Review, and Implementation Preparation

Timely introduction of programs into the market and smooth, uninterrupted operation of continuing programs requires sufficient time for program design/solicitation, selection/review, and preparation for implementation. (Implementation preparation includes such time-consuming activities as contracting, working out informal arrangements with trade allies, and developing program materials and processes.) This means that timing of these activities and the evaluation activities used to support policy decisions needs to be carefully planned if the process is to operate with optimum information, consistency, and efficiency.

The Value of Dispersed Timing for Evaluations

A large portfolio of programs launched with concurrent start and end dates can create large “spikes” in the workloads of various entities, including portfolio administration staff, contracting entities, program implementers, evaluation contractors, and other stakeholders. This can lead to inefficiency and, sometimes, efforts that have a higher probability for error due to strain on the resources available. If the program duration is fairly long, the lull between program evaluation contracts can create difficulties for evaluation firms who must manage their firms to match the ebb and flow of the evaluation market. These could discourage the most competent firms from remaining in the field, lowering the availability of high quality evaluators when they are needed.

Contract Requirements for “Pay for Performance” Programs

The contract requirements for “pay for performance” programs such as the Standard Performance Contract (SPC) program or other programs utilizing an Energy Services Company (ESCO) performance contract may require evaluation results on a time line different from the program or portfolio planning process. The same issue may also apply to incentives paid to utilities under a resource acquisition or procurement scenario. The program planning, implementation and evaluation cycle may not match the contracting requirements for payment levels based on ex-post or net evaluation results.

Market Inertia

Many information and market transformation programs take time to work due to significant market inertia. The timing of the effects evaluation needs to consider the likely response time of the market to the program intervention. Market effects evaluations may need to be conducted after multiple program cycles to provide enough time for the programs to overcome inertia in markets in which they operate.

Timing Needs for Retention, Measure Life, and Technical Degradation Analyses

Measure retention studies may need to be conducted at several points in time years after the program has ended, due to the dynamics of building remodeling, occupant turnover, operations and maintenance practices, business investment decisions, new technology entries into the market, and measure efficiency degradation. These studies need to be conducted periodically after there is a reasonable expectation of removal, failure, or degradation of a significant fraction of the measures.

Oversight and Review

Not only evaluation study plans, but also progress and reports may require some level of oversight and review. For example, evaluation study results may need to be reviewed for accuracy or their ability to support the portfolio planning process. Time to conduct these reviews should be included in the overall process.

Example of an Integrated Evaluation Timing System

The evaluation timing system described in this section explores an example of a method to integrate program planning, implementation, and evaluation into a continuous cycle of activities that provide timely feedback of information into the overall strategic planning process. The overall strategic planning process in this example has been divided into five major cyclic program activities, including:

- I. Goal Setting - Updating and Potential Analysis;
- II. Portfolio Analysis - Sector and Program Priorities, Public and Regulatory Review;

- III. Portfolio/Program Design, Selection, Public and Regulatory Review, and Approval;
- IV. Preparation for Implementation, Program Launch Preparation, Overall Evaluation Planning, and Regulatory Review; and
- V. Program Implementation, Evaluation, M&V, Market Assessment, and Ongoing Regulatory Oversight.

These activities and their possible placement in this example are discussed below.

Activity I: Goal Setting - Updating and Potential Analysis

Activity I consists of establishing the high-level goals for the efficiency portfolio and estimating the achievable potential for the efficiency resource. Overall energy efficiency policy goals (such as targeted percentage reduction in per capita energy use) are restated in terms of specific energy and demand reduction targets. Overall policy goals are reconciled with load forecasts and other information on utility energy and capacity requirements.

Once the specific energy and demand goals are defined, the potential analysis helps refine the goals based on the overall availability of the conservation resource within technologies, market sectors, and geographic area. Technology and market performance information from program evaluations, market saturation studies, and other overarching studies are incorporated into the efficiency resource potential estimates. Information is also assessed regarding the energy and demand goals, potential analysis, and other policy objectives, such as geographic and sector equity issues.

Activity II: Portfolio Analysis – Sector and Program Priorities, Public and Regulatory Review

A preferred portfolio of energy efficiency programs for the upcoming program cycle is defined based on the goals and potentials established in Activity I and further review of evaluation data on program impacts, process and market effects, and past overarching studies. Lessons learned from current and past program efforts along with efficiency potentials by technology and market sector from Activity I are used to define a portfolio of programs that meet the overall efficiency resource goals. The results of the portfolio analysis are presented for public comment and regulatory review. Preliminary evaluation planning and budgeting may also take place during this activity.

Activity III: Portfolio/Program Design, Selection, Public and Regulatory Review, and Approval

Designs for specific programs meeting the portfolio goals defined above are developed and/or solicited and selected. The proposed portfolio of programs is finalized and presented for public comment and/or approval by the responsible entity.

Activity IV: Preparation for Implementation, Program Launch Preparation, Overall Evaluation Planning, and Regulatory Review

Once the program portfolio is designed, selected, and approved, the programs are designed and structured for implementation. Program launch relationships, materials,

activities and timing strategies are finalized and made ready. Contracts (if needed) are negotiated, trade allies and key stakeholders are notified, and materials and internal processes are developed to prepare for program introduction and program launch efforts. Concurrently, an overall evaluation planning process takes place. This process establishes overall goals and objectives for the program evaluations, sets evaluation priorities based on perceived risks to achieving the portfolio savings objectives and other objectives, and addresses program design issues related to evaluation. Issues such as evaluation resource allocation, evaluation study consolidation, the need to evaluate a particular program during a given program cycle, and evaluation scheduling are addressed during the overall planning process. The overall evaluation plans will need to be reviewed by an appropriate body or bodies to assure that they will meet the information needs of policy makers, portfolio managers, program administrators, and program implementers. This examination considers not just the information needed about the planned program effects, but also the information needs associated with portfolio planning and technology targeting, including such tasks as updating the DEER and maintaining accurate persistence estimates.

Activity V: Program Implementation, Evaluation, M&V, Market Assessment, and Ongoing Regulatory Oversight

After the overall evaluation plans have been approved, more detailed evaluation plans must be developed and evaluation professionals selected to conduct the studies. The overall evaluation plan developed in Activity IV will likely schedule new or redesigned programs to initiate early process evaluation coordination and support activities as well as M&V planning during the initial program start-up period. These early actions can help assure that program designs support the evaluation function, in addition to helping establish the platform from which early evaluation feedback can be initiated. These early actions can also lead to improved database designs that help support the evaluation efforts. Following these early efforts, the evaluations will then likely swing into standard evaluation activities after the program is up and running, but early enough in the program cycle to provide feedback and corrective recommendations to program implementers in time for the program to benefit from those recommendations. Early impact evaluation activities to support program progress tracking consist primarily of measure installation tracking and verification combined with ex-ante savings estimates by measure. Adjustments to ex-ante savings estimates may be made based on early issues identified during M&V activities.

The full net impact evaluation analysis proceeds according to the schedule laid out in the program evaluation plan. Ex-post savings by measure and/or program are estimated, and the final program impacts are estimated based on the final program accomplishments. Assumptions underlying the efficiency potential analysis can then be updated based on the full net impact analysis. These data then feed back into the goal setting and potentials analysis activities, and the cycle repeats to allow for an integrated planning process. These activities are all conducted with an oversight process by an organization or organizations of appropriate responsibility.

The above example of a coordinated and integrated system is one example of how this process can be structured. This document makes no recommendations on the organizations responsible for these efforts, but presents this example of how a system might function to allow for a coordinated process in which the steps that need to be integrated are established along a single timeline.

The steps and feedback paths for this type of an integrated planning process are shown in Figure 5.1.

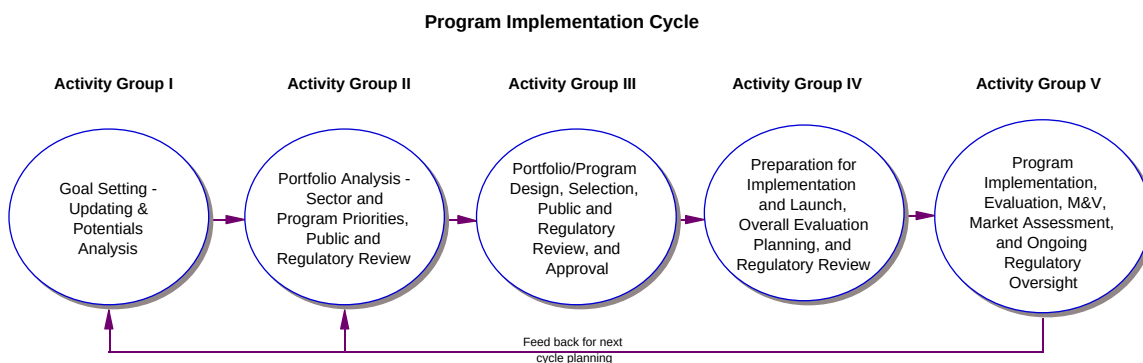


Figure 5.1: Program Implementation Cycle

Note: See the appendix to this chapter for a more detailed example of this process with key steps and dates incorporated into a sample timeline.

On top of this cycle of repeating activities, longer timescale analyses of market effects and measure persistence can be conducted to provide important information to the integrated planning process. A discussion of the issues and the timing of these efforts are described below.

Program Duration

Programs should be conducted over a period of time needed for the programs to establish themselves in the market and reach stable operations prior to conducting the program net impact evaluation. In addition, this period should allow for program improvements to be made to the programs, as a result of the evaluation recommendations and findings, while the program is still operational.

Timing of Evaluation Activities

Ex-post measure savings analysis and net-to-gross analysis should be completed prior to program completion to allow the evaluation information to feed back into the planning process for the next program cycle.

Multiple Offset Program Cycles

Separate program cycles with offset start dates can be considered to provide opportunities for new program proposals on a more frequent basis and to level the load for all parties involved in planning, selecting, developing, implementing, and evaluating programs. “Off-cycle” programs may be targeted to fill particular gaps in the portfolio and/or focus on innovative programs, pilot programs, demonstration programs, or emerging technologies.

The interactions between two separate program tracks, with a two-year offset between each track are shown in Figure 5.2. Note that the second track program planning cycle can take advantage of impact evaluation information from the first track in the overall planning process.

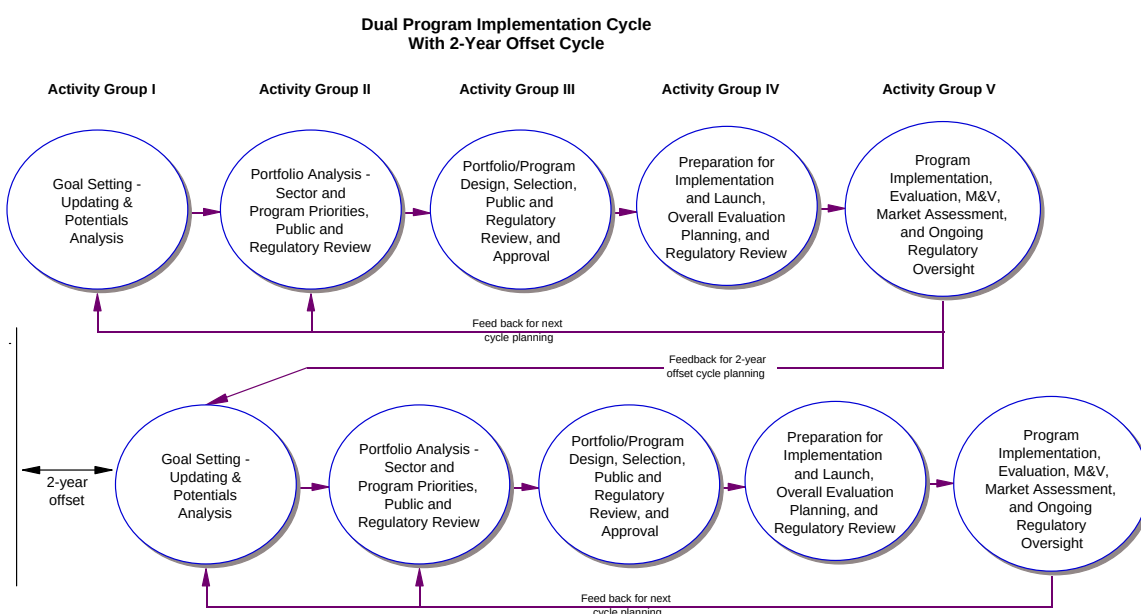


Figure 5.2: Two-Track Program Implementation Cycle

Impact Analysis for Final Contract Purposes

Final program results for the purpose of settling contract terms and determining incentive payments should be derived from the ex-post analysis and final program accomplishments shortly after program completion. These efforts can be held separate from the impact analysis conducted to inform the program planning process.

Program Planning Period

A coordinated effort between the evaluation, efficiency resource potential, and portfolio analyses will be required to complete the necessary analysis in time for it to impact the next integrated planning effort to guide the next cycle of programs. For programs that are

contracted, solicitation and contracting procedures may need to be streamlined to ensure that an integrated planning process can be accomplished.

Detailed examples of the timing of an integrated planning process are presented in the appendix to this chapter.

Roadmaps and the Evaluation Planning Process

This document contains several chapters specifically designed to support the evaluation planning process. Also, each chapter discussing a specific type of evaluation (impact, process, market effects, etc.) contains a planning decision guidance “roadmap” that allows the reader to see when and under what conditions a specific type of evaluation may be desired. These chapters provide valuable guidance concerning the CPUC’s expectations for the rigor, quality, and focus of the evaluation efforts and the associated activities needed to support the evaluation function.

In summary, the chapters presenting the evaluation roadmaps and decision systems cover the following evaluation topics:

- Impact evaluations;
- Measurement and verification efforts (M&V);
- Process evaluations;
- Market transformation evaluations;
- Evaluations of information and education programs;
- Non-energy effects evaluations;
- Sampling requirements and methods;
- Uncertainty and bias issues associated with the evaluation efforts; and
- Designing evaluations with cost-effectiveness analysis needs in mind.

The information presented in these chapters is provided to help guide the evaluation planning and implementation process for California’s energy efficiency programs. These chapters will help the evaluation planner make decisions about, and develop the methods and procedures for each evaluation. These chapters will help guide the evaluation planning process and the activities associated with conducting the field efforts, analyzing the data collected, and in developing the evaluation reports.

The program-specific evaluation planning process should move through each type of the program evaluations presented in this document according to the type of program being implemented. These efforts should be planned to take place within the contracted program implementation period so that the evaluation results can be used to help improve the program to the extent possible within the implementation period, to document program accomplishments and net benefits, and to inform the portfolio assessment of which programs to offer in the future. The evaluation planning and implementation process should match the evaluation needs associated with the goals of the program. This means different types of programs should plan for different types of evaluations efforts.

These efforts and the key policy and regulatory issues associated with each of the types of programs and program evaluations are discussed below.

All Programs

There are key parts of the Framework designed to feed the evaluation planning process for all types of programs and all types of evaluations. These chapters focus on sampling methods and approaches to support the evaluation effort, on having the evaluation planning process focus on, and address, uncertainty and bias issues associated with each evaluation effort, and on designing evaluations that inform the program-level and the portfolio-level cost-effectiveness analysis needs.

The key policy-related, regulatory and evaluation plan review and approval issues regarding these chapters include:

- Is the sampling plan representative of the population served?
- Is the sampling plan able to support the evaluation policy objectives?
- Is the sampling plan objective and unbiased focusing on key evaluation needs?
- If the sampling plan is biased, how can it be improved?
- If the sampling plan is biased, how do the biases harm the evaluation's objectivity or influence our understanding of the program's effects?
- Are there threats to the validity of the evaluation results that are incorporated into the evaluation design?
- Can the evaluation design be modified to limit threats to the validity of the evaluation results?
- Have or will the instruments proposed be examined for potential threats to validity? Is there an estimate of the magnitude of these effects or to the degree that these are occurring? Are methods proposed to address these effects? Are cost justified mitigation methods proposed?
- Are the threats to validity or biases described in the evaluation plan so that policy makers understand these threats or biases and the limitation these provide to the use and usefulness of the evaluation data and to the findings developed from the evaluation?
- Does the evaluation provide the data needed to inform a cost-effectiveness assessment of the program and to compare this program with others?

Programs with Impact Goals

Programs with energy impact goals should focus their evaluation plans to include an energy impact evaluation and a process evaluation over the program's implementation cycle. Evaluation planners for these programs will want to read the Framework chapters on impact evaluations, on conducting measurement and verification (M&V) efforts, and on conducting process evaluations.

The key policy-related, regulatory and evaluation plan review and approval issues regarding programs with impact goals include:

Impact Evaluations

- If no impact evaluation is proposed, is this well-justified based upon documentation of the following:
 - Prior net impact evaluation history for this program, date of last impact evaluation, and stability/precision of results,
 - Comparability of program prior net energy impacts with other similar programs,
 - Size of program impacts and program budget and current evaluation budget limitations,
 - Maturation of the program, and
 - Similarity of program with past programs for which net impact evaluation results are available (technology mix, delivery mechanism, customer mix).
- Is the evaluation using appropriate analysis methods considering the type, size and timeline of the program being evaluated and for the data that is available or can be collected? (See Framework Appendix C for an example assessment.)
- Will the evaluation reliably report gross and net impacts associated with this program?
- Will the evaluation reliably report the gross and net impacts associated with the program and program technologies and can these be used to update the DEER for impact, cost and measure life?
- Are the verification efforts adequate to document service and to support impact assessments?
- Are the M&V efforts designed to support the net impact assessment needs so that impacts are accurate for this program, as implemented?
- Are the impact and M&V methods and approaches appropriate to support the policy formation and regulatory needs?
- Will the evaluation test the program theory's causal relationships and assumptions regarding program events and activities, and program results so that the theory can be supported or modified?

Process Evaluations

- If no process evaluation is proposed, is this well-justified based upon documentation of the following:
 - Prior process evaluation history for this program, satisfactory or good process evaluation findings within last three program cycles,
 - Program has not changed since last process evaluation was conducted,
 - Size of program impacts compared with the program budget and the current evaluation budget,
 - Maturation of the program,
 - Similarity of prior process evaluation results for past programs for which process evaluation results are available, and
 - Program impact and cost-effectiveness results have shown to be both close to expectations and similar to other programs offering similar services.

- Is the process evaluation designed to be able to identify weakness in the program's designs, methods of operation, or implementation procedures that limit the ability of the program to obtain cost-effective impacts?
- Will the process evaluation identify program methods and procedures that can be modified to increase energy impacts, improve customer satisfaction, increase enrollment rates, or improve market appeal?
- Will the process evaluation document the program's methods of operation so that they are well understood by policy makers?

Information and Education Programs

The information and education program evaluation plan should contain a program effects evaluation and a process evaluation over the program cycle (unless they have energy impact or market transformation goals, see chapter 5 for details). Evaluation planners for these programs will want to read the Framework chapters on evaluating information and educational programs and on conducting process evaluations.

The key policy-related, regulatory and evaluation plan review and approval issues regarding programs with information or educational goals include:

Program Effects Evaluations

- Will the evaluation document and, to the extent possible, quantify the direct effects of the program on participants or the recipients of the program's services?
- Will the evaluation test the program theory's causal relationships and assumptions regarding program events and activities, and program results so that the theory can be supported or modified?

Process Evaluations

- Is the process evaluation designed to be able to identify improvements in the program's methods of operation or implementation procedures that can improve the ability of the program to accomplish its informational or educational goals or increase the intended effects from these efforts?
- Is the process evaluation designed in a way to obtain the information to allow it to identify new methods and procedures, or identify changes to current methods and procedures that could increase knowledge gained, the extent of the information used, or other improvements that would help the program more cost-effectively achieve its effects goals?
- Is the process evaluation designed such that it can obtain the information and identify and recommend changes to the program that might increase the level of desired behaviors?
- Will the process evaluation document the program's methods of operation so that they are well understood by policy makers?

Market Effects or Market Transformation Programs

Market transformation programs should plan for a baseline study or a baseline analysis approach that can serve as an evaluation baseline, a process evaluation and, at some point in time after which market effects can be expected, a market effects evaluation.

Evaluation planners for these programs will want to read the Framework chapters on market transformation evaluations and on conducting process evaluations as well as the 2001 publication on evaluating market effects and market transformation programs.³⁸

The key policy-related, regulatory and evaluation plan review and approval issues regarding programs with market transformation goals include:

Market Effects Evaluations

- Does the baseline study used to support the market effects evaluation provide an adequate presentation of the way the market operates relative to the markets that are targeted?
- Is the baseline information based on research that is conducted with enough reliability that the error bands around the baseline market data can be compared with the results from the market effects evaluation and used to determine market movement?
- Is the baseline study capable of establishing a measurable baseline from which market effects metrics influenced by the program can be tracked and measured for program effects?
- Will the evaluation reliably (to the extent possible) document the net changes that have occurred in the market as a result of the program?
- Does the evaluation have a reasonable approach for disaggregating market effects caused by the program with the market effects caused by non-program events so that net market effects can be estimated?
- Does the market effects evaluation have a reasonable approach for forecasting how the current market effects will continue in the future?
- Will the evaluation test the program theory's causal relationships and assumptions regarding program events and activities, and program results so that the theory can be supported or modified?

Process Evaluations

- Is the process evaluation designed to be able to identify weakness in the program's designs, methods of operation, or implementation procedures that limit the ability of the program to change the way markets operate to be more energy efficient?
- Will the process evaluation identify program methods and procedures that can be modified to increase market effects?
- Will the process evaluation document the program's methods of operation so that they are well-understood by policy makers?

³⁸ See *A Framework for Planning and Assessing Publicly Funded Energy Efficiency*. (* Sebold et al. 2001).

Programs That Rely On Non-Energy Effects to Achieve Their Goals

Some programs may rely on non-energy effects to gain participation and achieve their energy savings goals. For these programs there may be a desire to conduct a non-energy effects evaluation. These evaluations can be approved for funding if they pass specific conditions regarding the need and use for the evaluation findings. Evaluation planners for these programs will want to read the Framework chapters on conducting evaluations on non-energy effects.

The key policy-related, regulatory and evaluation plan review and approval issues regarding programs with non-energy effects impacts include:

- Is the non-energy effects evaluation needed to more accurately understand or to increase the energy effects of the program?
- Does the CPUC have a policy supporting the spending of Public Goods Charge funds to document the specific non-energy effects targeted?
- Can the evaluation design reliably identify and quantify the targeted net non-energy effects?

Overview of the Framework

This document and the associated roadmaps provide a structured approach to planning and conducting evaluations of California's energy programs. Figure 5.3 provides an overview of the process and allows the reader to see the components of the evaluation framework structure and linkages between the components. These steps and processes are presented and discussed in detailed in the associated chapters of this document.

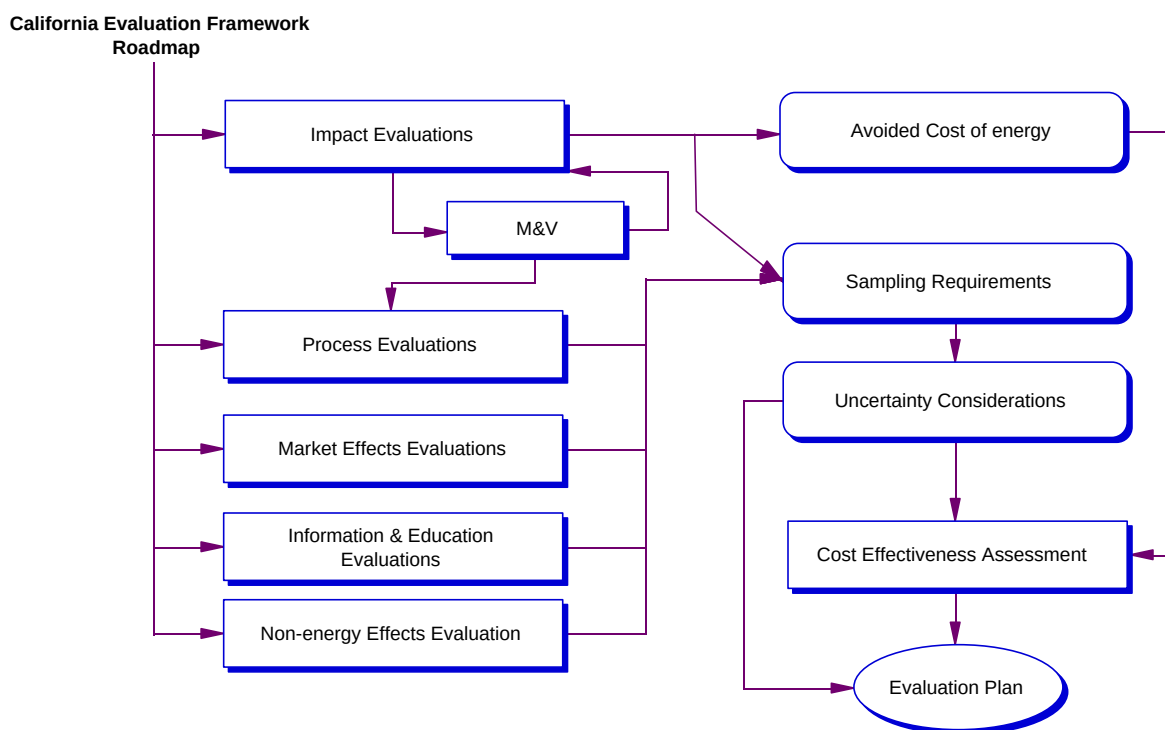


Figure 5.3: Overview of the Evaluation Framework

As discussed earlier, the Framework is an evaluation guidance document and is not a prescriptive evaluation protocol. While the Framework methods and approaches are not mandatory, they are developed to help guide the evaluation planning process for California's energy efficiency programs. This document is intended to help guide the evaluation process to provide reliable and comparable evaluation results across the different types and sizes of programs, and to focus on and address specific issues related to the evaluation planning and implementation process.

Determining Evaluation Priorities

Evaluation funds are generally limited to a small fraction of the total program budget and this small fraction may not be adequate to provide all the evaluation results needed to make well-informed policy or program decisions. Therefore, there is almost always a need to carefully allocate evaluation funds among the various potential evaluation activities that must be pursued. This frequently raises a question of what relative emphasis to place on each of the two key evaluation functions (i.e., documenting and measuring program effects versus understanding/improving program processes) and on each of the types of evaluation, as well as the question of which particular programs should receive the most emphasis.

There is a discussion of a systematic approach for allocation of evaluation resources in Chapter 12, Uncertainty. This approach is recommended, at least for the consideration of resource allocations between various program impact evaluations.

A program-level allocation of evaluation resources should be examined when program-level budgets are set. However, there are also many qualitative decisions that need to be made, or decisions that are difficult to quantify.

The decision on how to allocate evaluation resources within a program-specific evaluation budget should be made after a careful consideration of the specific program and the policy decisions being confronted. Some examples of key factors that should be considered include:

- What are the most important information needs of the key policymakers/regulators associated with the program?
- Where are the programs in terms of their implementation cycle (e.g., “new” programs generally need a relatively higher level of process evaluation emphasis than existing, well-established programs)?
- What key decision points are approaching and on what time schedule?
- What level of evaluation information already exists and where are there important information gaps?

In short, there are no hard and fast rules about relative allocations of evaluation resources to various functions and types of evaluation. Those decisions must be made on a very situation-specific basis. However, there are several key information needs that represent the main purposes for conducting evaluations in California. Evaluation resources should be focused so that these needs are satisfied. These include:

- The need for accurate net program-level energy and demand impact information that reliably indicates the resources gained as a result of spending public dollars for the program funded,
- The need to understand the net energy and demand impacts associated with the individual technologies and behaviors adopted, as implemented and used,
- The need to accurately compare the cost-effectiveness of the program with other programs making up the portfolio of programs,
- The need to understand and have measurements of the program’s key effectiveness indicators,
- The need to understand the program theory supporting the structure, organization, operations and activities of the program and to test the program theory and the theory’s underlying assumptions,
- The need to understand the market and the key market factors effecting the program’s ability to achieve energy and demand goals and achieve the program’s operational and market-related goals,
- The need for an understanding of the program and its operational characteristics and methods to identify those that can be improved, and to understand those that should be used by other programs,
- The need for accurate information on the non-energy program goals achieved by the program and the effects of those achievements,

- The need to understand if the program has reached its best cost-effective performance or if it can be expected to improve its cost-effectiveness through program design, operational, or implementation changes,
- The need for information that helps policy makers determine the ability of the program to produce energy and demand savings in the future,
- The need to document the activities and results received in exchange for Public Goods Charge or procurement funds, and
- The need for information to help program managers understand the results of their program's efforts and to provide timely feedback to managers to help them improve their programs.

These above needs represent the primary reasons for conducting evaluations in California and provide guidance to the evaluation planning process. These needs are demanding on the evaluation effort and on the resources that can be allocated to the evaluation function. In some cases these needs may not be met within a single program implementation and evaluation cycle. As a result, the evaluation planning efforts may need to address the timeline over which the needs can be addressed, and focus the current program-cycle evaluation efforts on those needs that are identified as the highest priority for the current cycle.

Recommendations for Setting Evaluation Budgets

The budgeting of evaluations is a subject of interest to all stakeholders in the program planning, implementation and evaluation cycle. Estimates of where evaluation budgets should be set have in the past generally ranged from about ten percent to a low of about two percent of the program budget. Budgets for the 2002 – 2003 programs averaged just over four percent. A great many of these evaluation plans indicated that the lack of evaluation resources significantly restricted the ability to reliably estimate program effects, especially at the technology level. Many of the 2003 – 2004 program evaluations relied on deemed savings estimates or adjustments to deemed savings. In addition, many program administrators indicated that they had insufficient resources to conduct process evaluations or to obtain needed baseline information for their programs.

The most important consideration for establishing evaluation budgets is the need for reliable estimates of energy effects and program potentials to be used for making energy resource policy decisions and for identifying cost-effective energy supply options. The second most important reason for evaluation is to improve the operations and cost-effectiveness of energy programs. Evaluation budgets that are set too low to accomplish these goals are of little value to any stakeholder. Evaluation budgets that are set higher than needed for these goals may be a waste of valuable resources.

There is no single specific percentage of a program's budget that should be allocated for the independent evaluation process. Evaluation budgets for small pilot programs, testing new designs or delivery concepts, may need to be set at a level higher than the program costs in order to collect and analyze the information needed to determine if the program

should be continued or expanded. In these cases it would not be unusual for a pilot test program to have a program budget of \$100,000 and an evaluation budget of \$100,000. At the other end of the scale are large statewide programs offering technologies and services that have been evaluated in the past, offered in well-understood environments in which the energy impacts and operational procedures are well-documented, and in which the procedures are working smoothly and efficiently. In this case a \$15,000,000 program may need only two percent of the program budget to verify what is already understood. Most energy programs can be pegged somewhere between these two extremes.

One of the key questions that needs to be addressed in establishing a program-level evaluation budget is found in the answer to the question: What are the energy supply decisions that must be made in relation to the performance of this program and what is the supply risk of making a wrong decision relative to this program and these technologies? This is a decision for policy makers and for portfolio managers charged with providing an overall portfolio of low-cost and reliable energy supply and demand-side resources.

In establishing an evaluation budget policy, one must also look at the cost of the program and the number of years over which the program may be implemented. Some may suggest that a \$100,000 program is too small to worry about conducting a rigorous or reliable evaluation. Yet if that program is funded for a period of ten years, the program becomes a \$1,000,000 program that is never evaluated. Policy makers are then placed in a position to defend the spending of a million dollars without an understanding of the program's effects or potential effects.

At the end of the day, and after all arguments have been aired, policy makers still need to allocate a budget to the evaluation effort. Adequate budgets are generally lower than ten percent but higher than four percent. Budgets can stay toward the lower end when prior evaluations are drawn upon and circumstances are explored to determine the necessity for repeat evaluations.

In order to establish an evaluation framework that incorporates a wide range of evaluation needs into a program cycle evaluation planning process, the evaluation budget needs to be in the range of six to eight percent of the total portfolio budget. However, this budget should not be equivalently distributed across the programs. Rather, the budget should be allocated at the program level based on the need for reliable information after careful examination of past program evaluations, an assessment of the rigor (reliability of findings and levels of uncertainty) of those evaluations, an understanding of the current mix of programs and technologies, and their associated mix of targeting and delivery practices. This approach will help assure that evaluation resources are wisely used and that the information gained from the evaluation improves the ability of policy makers and energy supply planners to make informed decisions.

Evaluation Spending Priorities

While it is important to understand the budgets that are typically considered appropriate for energy efficiency programs, there are evaluation priorities that can serve to inform decisions about how those funds are distributed among the various evaluation goals.

1. As a general rule of thumb, evaluation efforts for resource acquisition programs first need to focus on establishing reliable net energy impact estimates for both the program as a whole and for the technologies addressed in the program. However, it is also important to understand the key factors that drive net savings at the technology level for different types of participants. For information and education programs, it is important to understand the effects that are being achieved by these programs, to be able to identify what information or education efforts provide these effects and the causal relationships between the two. For programs that are designed to change how markets operate, the evaluation should first focus on being able to identify, monitor and track the net changes in the market being caused by these programs and the causal linkages between program actions and market effects. In addition to knowing the impacts/effects from program efforts, it is important to know that the investments are wise investments. This means that providing the information to assess the program's cost-effectiveness and cost-efficiency of the program also needs to be a top priority of the evaluation focus.
2. Maintaining high participant satisfaction scores and reducing barriers to participation represents the second tier of efforts on which evaluation resources can focus. California is dedicated to providing customer-focused services that are designed to achieve high levels of customer satisfaction. To maintain these levels, it is necessary to understand the conditions that drive participant satisfaction and identify methods for increasing customer satisfaction scores. California energy efficiency programs should continually strive for the highest levels of participant satisfaction while obtaining the highest level of net benefits (cost-effectively). To accomplish this goal it is important to understand how energy efficiency programs, products and services need to be structured and delivered to achieve cost-effective energy resources or achieving other program goals while maintaining high participant satisfaction. Additionally, programs must be aware of the barriers that keep customers from becoming participants and design programs that are able to reach and enroll eligible participants within the market segments that are targeted.
3. The third tier of evaluation efforts focus on identifying ways to improve the cost-effectiveness and operational efficiencies of the programs in order to evolve the program toward its most cost-effective or resource efficient mode of operation. These efforts can focus on a wide range of activities but typically involve the examination of the program designs and methods of operation and included investigative activities that focus on improving a wide range of program

characteristics, including; program design, management, administration, staffing, targeting, marketing, processing, service delivery, reporting, and other efforts.

4. Finally, program evaluation resources need to document what the program is accomplishing in exchange for the public dollars spent. In the end, all energy efficiency programs funded by the Public Goods Charge and procurement dollars are accountable to the people of the state of California. The documentation function associated with providing energy efficiency programs is provided by several functions, including administrative tracking and program progress reporting. However, in addition, it is the evaluation studies and reports that can serve as the independent source of information on what is provided and what is achieved as a result.

Consolidating Programs for Evaluation

The process of grouping programs together or consolidating the evaluations of multiple programs may provide opportunities to increase the efficiency of the evaluation effort or the ability of the evaluation to produce more reliable results (e.g., decrease uncertainty levels). Coordination or consolidation options that may improve evaluation efficiencies include:

- Overlapping instrument and methodology development to serve multiple programs;
- Overlapping or complementary sample selection (which may increase ability to obtain sample points for specific technologies or types of customers/facilities, etc.);
- Overlapping field research or data collection efforts; and
- Data sharing or analysis grouping to increase reliability of findings.

In addition, consolidating of the evaluation function can, in some cases, increase the level of interest in the project from firms that may not typically bid on small evaluation projects, thereby increasing the level of competition for the study.

Consolidating the evaluations of similar types of programs or consolidating evaluations by their evaluation efforts has significant potential for increasing evaluation efficiencies as well as for providing additional depth, rigor of measurement, comparability, and value from the study. For example, it is possible to consolidate impact studies across different types of programs. Likewise, process evaluations can be consolidated, especially when they are testing similar types of program designs, operations or systems. Field metering and verification efforts can also be consolidated across a wide range of program types.

Consolidation of the evaluation effort can also be aggregated by market sector. For example, a group of residential programs can be consolidated. Likewise, a group evaluation of programs focusing on industrial or agricultural markets may be practical.

Evaluation consolidation may also be structured by program size, especially for small programs. Small programs are often resistant to spending evaluation dollars that may be needed to keep the program running. However, these programs may be consolidated to obtain an evaluation budget that can address a portion of the evaluation needs. In this case, it may also be possible to structure the evaluation over multiple program implementation cycles and focus each cycle on a set of evaluation objectives.

Finally, another approach to consolidation would be for multiple programs that are offered by the same implementer, so that the program evaluation efforts, across multiple programs, would deal with the same organizational structure. The value of this type of consolidation is less certain, because the evaluation contractor or contractor teams may need to demonstrate expertise across a wider range of program types, technologies, and evaluation approaches with relatively little overlap.

Consolidating the evaluation functions across multiple programs can also add value to the evaluation effort. Examples of increasing the value of the evaluation via consolidation include:

- Information and educational programs targeting schools and teachers implemented by different organizations, perhaps using different strategies and approaches that can be compared;
- HVAC tune-up programs offered across the state so that the same or similar services can be evaluated across different markets, different administrators, different size units, or different targeting and delivery approaches;
- Multi-family programs that must overcome similar barriers but may do so in different ways so that different approaches, incentive levels, or levels of services can be compared;
- Programs that target mobile homes but have a mix of covered technologies or service designs that can be assessed relative to other programs serving the same market; and
- Small commercial lighting programs that have a similar mix of technologies but may focus on different sectors within the commercial market or may have different approach and delivery strategies that can be compared.

While there are often benefits associated with consolidating evaluation efforts across multiple programs, it is not always advisable. Despite its promise, consolidation often faces obstacles. These may include:

- Programs that are at different points in their implementation cycle, reducing some of the potential benefits of consolidation;
- Grouping evaluations into too large a group so that only a limited number of evaluation firms can handle the load at a given point in time, thereby restricting competition and giving advantage to large firms;

- Grouping evaluation into too large a group so that the study requires small evaluation firms to group and subcontract pieces of the work, coordinating across complex teams of evaluation professionals;
- Timing issues in which the evaluation funds are scheduled to be available at different points in time reducing the potential for consolidated efforts;
- Ability of the evaluation plan to be developed and approved across different programs within the timeline allowed for these efforts;
- Implementers' concerns over loss of focus of the evaluation on their program;
- Implementers' concerns over the potential for unfavorable or inappropriate comparisons;
- Programs with characteristics that need to be evaluated by firms with special expertise, indicating that an evaluation may need a different evaluation team than the other programs; and
- Programs that have been evaluated in the past for which special research approaches or methods have been developed that may not apply to other programs.

Appendix to Chapter 5

An Example of an Integrated Planning Cycle

This appendix provides two examples of how the program design, implementation and evaluation cycles can be structured to provide an integrated planning approach.

Example 1 – Three-Year Program Cycle with Integrated Planning, Solicitation, and Evaluation Process

An example of a three-year program cycle with an integrated planning, solicitation, and evaluation process is shown in this example. An example describing the individual steps within the five activities described in Chapter 5 is as follows.

Activity I. Goal Setting – Updating and Potential Analysis

Activity I consists of two steps. These steps make up the process by which the high-level energy efficiency portfolio goals are established and updated, based on overall policy objectives and analyses of the maximum achievable energy efficiency resource. Goals and resource potential baselines, once established, will undergo periodic evaluation, review, and updating to provide timely policy information related to the accomplishments of the portfolio of energy efficiency programs.

1. **Goal setting and updating.** In this step, the high-level goals for the energy efficiency portfolio are translated into specific energy savings and demand reduction goals. For example, an overall goal of a one percent annual reduction in per capita energy use within each IOU service territory are translated into specific MWh/MW numbers and targets. These targets, in turn, provide the necessary guidance to the Portfolio Analyses activities taking place in Activity II.
2. **Potentials analyses, baselines, and updates.** In this step, the available data on the potential for the energy efficiency resource are reviewed and a benchmark of achievable potential is established. The benchmark for achievable energy efficiency provides the information needed to refine the portfolio savings goals based on the overall availability of the resource, and its availability within specific sectors. This step also includes analyses of load forecasts and other information related to utility capacity and energy needs. This information is analyzed and used to determine the best combination of efficiency resources, sectors, and technologies for the Portfolio/Program Design activity, undertaken in Activity III.

Activity II. Portfolio Analysis – Sector and Program Priorities, Public and Regulatory Review

This activity has one step, as described below.

3. **Program, technology, and sector priorities.** Based on overall efficiency resource goals and the specific energy and demand targets identified in Activity I, program impacts analyses and evaluations, process and market effects evaluations, as well as other completed studies will be reviewed to design a preferred portfolio of programs for the upcoming program implementation cycle. This activity involves a thorough analysis of the status and success of the current and past energy efficiency portfolio efforts. The programs, market sectors and technologies identified in step three are those that will be included in the program portfolio. This portfolio, in turn, is expected to provide the energy and demand savings reductions necessary for meeting established savings targets. Once the preferred portfolio of programs is selected, the portfolio can undergo a review and approval process by a responsible entity, including, if desired, a public review and comment process.

Activity III. Program/Portfolio Design, Selection, Review and Approval

Activity three consists of three steps. These steps make up the process by which the portfolio is designed and approved, and programs are selected. This activity includes: selecting and/or soliciting programs and program portfolios identified in step three that are expected to acquire the resources identified in step one, obtaining the necessary regulatory approval of the portfolio design, and reviewing and selecting proposals received for competitively-bid programs.

4. **Program and portfolio design.** A portfolio of programs will be designed based on the results of the portfolio analysis conducted in Activity II. The portfolio design will address issues such as program types, technologies, markets, and delivery mechanisms, as well as issues such as equity and risk necessary to create a balanced portfolio.
5. **Program and portfolio selection.** Programs will be selected and/or solicited to meet the criteria developed in the portfolio design. Program selections consider the policy goals and energy savings targets specified in the portfolio design, as well as the abilities of the program or portfolio providers to accomplish the goals.
6. **Portfolio review and approval.** The preferred portfolio of programs is presented for review and approval by the responsible entity.

Activity IV. Preparation for Implementation, Program Launch Preparation, Overall Evaluation Planning, and Regulatory Review

Activity IV consists of several steps conducted in parallel involving program preparations prior to program launch, overall evaluation planning and review of the overall evaluation plan.

7. **Program implementation and launch preparation.** Detailed program design and implementation plans are finalized and the program is made ready to place in the market. Activities may include finalizing program theories to fit the way the market operates, testing program implementation strategies, finalizing program designs, contract negotiations with program implementers, trade ally and key

stakeholder communications, developing marketing materials, and finalizing launch sequencing and timing strategies.

8. **Overall evaluation planning.** Goals and objectives for conducting program level evaluations are established. Evaluation resources priorities are set and evaluation resources are allocated according to these priorities. Issues such as evaluation study consolidation, the need to evaluate a particular program during a given program cycle, and evaluation scheduling are addressed during the overall planning process. An overall plan for evaluating the programs within the portfolio is developed. Evaluation contractor selection criteria are included in the overall evaluation plan.
9. **Regulatory review.** The overall evaluation plan is reviewed by a responsible entity to assure that it meets the information needs of policy makers, portfolio managers, program administrators, and program implementers.

Activity V. Program Implementation, Evaluation, M&V, Market Assessment, and Ongoing Regulatory Oversight

Activity five includes the combined efforts for the program evaluation function, and covers a wide range of steps, from the selection of the evaluation contractor to the completion of the various types of evaluation studies.

10. **Evaluation contractor selection.** Evaluation firms are selected to conduct each program evaluation designated in the overall evaluation plan according to criteria contained within the overall evaluation plan.
11. **Develop detailed program-level evaluation plans.** Evaluation firms selected to conduct the program evaluations develop detailed evaluation plans for each program based on the goals and objectives specified in the overall evaluation plan.
12. **Process evaluations.** Process evaluations consistent with this Framework document are conducted. These efforts will generally take place after the program is allowed to develop and implement its operational and implementation procedures, typically six months after program initiation.
13. **Measurement and verification.** Data collection activities preceding the full net impact evaluation are conducted, which generally includes measure installation verification, field data collection, on-site surveys, metering and monitoring, billing data collection and preparation, and other such activities. M&V activities are coordinated with the process evaluation to provide early feedback to program implementers. These efforts are conducted early enough to inform the impact evaluation efforts, but late enough that the program is successful at gaining participants and is installing or getting ready to install program measures.
14. **Impact evaluations for first program year.** First year impacts are estimated based on measure installation verification and ex-ante savings estimates.

Measurement and Verification efforts are in progress and may be used to perform a preliminary adjustment of the ex-ante savings estimates based on early observations of measure installation quality and field performance. This process will make use of the established program progress tracking system.

15. **Impact evaluation for second program year.** The second program year impact evaluation consists of the use of the measurement and verification results along with the more rigorous net effects impact evaluations. Net program impacts for this program cycle are estimated according to this Framework. Data collection and analysis covering the first two years of program operation are used to inform this process. The impact evaluations provide the load impact information on the schedule required to conduct a portfolio analysis for the next cycle of programs. Ex-post energy savings data are developed to inform the third year end of cycle true up (below at step 16), provide revised savings estimates for program continuation proposals, provide information for consideration in DEER³⁹ revisions, and inform the market potential analysis for the next cycle program portfolio.
16. **Impact evaluation for third program year.** The third year impact evaluation consists, in general, of confirming the net impact analysis conducted in the second program year and includes a review of the technologies installed and their use conditions in order to “true up” the second year assessment into more reliable third year net impacts. Ex-post energy savings and net-to-gross estimates developed from the second year full net impact evaluation above are combined with third year program accomplishments to estimate third year program impacts. Programs receiving funding for an additional three-year cycle may also include third year activities with first and second year activities in the subsequent program cycle to perform the full net impact evaluation covering multiple program cycles.
17. **Annual update of program impacts vs. program goals.** In some cases, there may be a need to update program impact estimates and compare those estimates with program goals to determine if the programs are providing energy and demand impacts consistent with contractual requirements. This effort is done annually, using the best available impact evaluation results at the time of the update.
18. **Market effects baseline studies (market transformation and information programs).** Baseline studies are conducted to understand the relevant baseline characteristics of the markets prior to the introduction of market transformation and/or information programs.
19. **Market effects studies (market transformation and information programs).** Program evaluation efforts for market transformation programs or energy information and educational programs are conducted. In this step the evaluation

³⁹ DEER: Database for Energy Efficient Resources. (Xenergy et al. 2001). See <http://www.energy.ca.gov/deer/> for information.

focuses on the changes to the market that are caused by the programs. These evaluations are not energy impact studies, but studies that document the changes made to the market as a result of the program. These studies also examine the effect of information programs that are designed to change the way customers use energy, or act as promotional efforts to channel participants into other programs that directly deliver energy impacts.

- 20. Individual program future potentials estimates.** This step is an assessment of the future energy efficiency or demand reduction potentials associated with the program based on the impact, process and in some cases the market effects evaluations. In this assessment the evaluation teams examine the impact and available process and market effects information as well as the latest energy efficiency potentials study results and makes an assessment of what they consider to be the potential for the program under review over a longer period of time. This information is used to compare with the statewide market potential assessments and used to guide the setting of energy and demand targets for the next cycle of programs undertaken in Activity I.
- 21. Optional longer term impact & persistence studies.** Because the three-year process is cyclic and expected to reoccur such that there is a steady stream of energy programs and program portfolios, an opportunity exists to schedule longer-term impact evaluations during the following program cycles that return to the participants of the previous cycles to assess the degree to which the impacts are still being provided, and the persistence of the measures that are providing the impacts. These longer-term follow-up cycle impact studies are used to true up the impacts of the program or program portfolios over a period of time longer than a single program cycle. This provides a method to look at six, nine, and twelve year effects and to document the longer-term reliability of the energy resources from California's energy and demand programs.

A detailed schedule for this example, assuming a January 2006 program start date, is shown in Table 5.1 below.

Table 5.1: Example 1 Schedule

Activity	Step	Description	Start Date	End Date	Duration (mo)
I		Goals Setting and Potentials Analyses			
	1	Goals setting and updating	11/1/2004	2/1/2005	3
	2	Potentials analyses, baselines and updates***	8/1/2007	2/1/2008	6
II		Portfolio Analyses and Assessment			
	3	Program, technology and sector priorities	2/1/2005	4/1/2005	2
III		Portfolio/Program Design, Review and Selection			
	4	Program & portfolio design	4/1/2005	6/1/2005	2
	5	Program & portfolio selection	6/1/2008	8/1/2005	2
	6	Portfolio review and approval	8/1/2005	10/1/2005	2

IV		Preparation for Implementation, Program Launch Preparation, Overall Evaluation Planning, and Regulatory Review			
	7	Final design strategies and launch preparation	10/1/2005	1/1/2006	2
	8	Overall evaluation planning	10/1/2005	11/1/2005	2
	9	Regulatory review	11/1/2005	1/1/2006	1

Table 5.1: Continued

Activity	Step	Description	Start Date	End Date	Duration (mo)
V		Program Implementation, Evaluation, M&V, Market Assessment and Ongoing Regulatory Oversight			
		Program start-up	1/1/06	1/7/06	6
	10	Evaluation contractor selection	1/1/2006	4/1/2006	3
	11	Develop detailed program-level evaluation plans	1/1/2006	4/1/2006	3
	12	Process evaluations	7/1/2006	7/1/2007	12
	13	Measurement and verification	9/1/2006	7/1/2007	10
	14	Impact evaluation for first program year	1/1/2007	2/1/2007	10
	15	Impact evaluation for second program year ^{***}	1/1/2007	1/1/2008	12
	16	Impact evaluation for third program year	11/1/2008	3/1/2009	4
	17	Annual update of energy impacts vs. goals	1/1/2007	3/1/2007	2
			1/1/2008	3/1/2008	2
			1/1/2009	3/1/2009	2
	18	Market effects baseline studies (MT & Info programs)	7/1/2006	11/1/2006	4
	19	Market effects studies (MT & Info programs)	7/1/2007	1/1/2008	6
	20	Individual program future potentials	9/1/2007	12/1/2007	3

^{***} Second year impact results feed step 2 in subsequent cycle.

A Gantt chart showing the schedule for each of the steps in the three-year program cycle described in this example is shown in Figure 5.4.



Example 2 – Offset Program Cycles

This example shows two overlapping tracks of three-year program cycles that allow for more frequent program solicitations and portfolio updates. In this example, the schedule for the second program track will start two years after the first program track.

The example presents the schedule of activities resulting from operating two different program tracks, and explores the resource requirements needed to respond to the assessment, contracting, implementation and evaluation efforts associated with these activities. The schedule presented in the previous example is merged with the schedule for the second program track in Table 5.2.

Table 5.2: Schedule of Activities for Two-Track Example

Assumes a January 2006 start date for Track 1 and a January 2008 start date for Track 2
Track 2 activities are shown in italics

Track	Activity	Step	Description	Start Date	End Date
1	I	1	Goals setting and updating	11/1/2004	2/1/2005
1	II	3	Program, technology and sector priorities	2/1/2005	4/1/2005
1	III	4	Program & portfolio design	4/1/2005	6/1/2005
1	III	6	Portfolio review and approval	8/1/2005	10/1/2005
1	IV	7	Program launch preparation	10/1/2005	1/1/2006
1	IV	8	Overall evaluation planning	10/1/2005	11/1/2005
1	IV	9	Regulatory review	11/1/2005	1/1/2006
1	V		Program start-up	1/1/2006	7/1/2006
1	V	10	Evaluation contractor selection	1/1/2006	4/1/2006
1	V	11	Develop detailed program-level evaluation plans	1/1/2006	4/1/2006
1	V	12	Process evaluations	7/1/2006	7/1/2007
1	V	18	Market effects baseline studies (MT & Info programs)	7/1/2006	11/1/2006
1	V	13	Measurement and verification	9/1/2006	7/1/2007
2	<i>I</i>	<i>1</i>	<i>Goals setting and updating</i>	<i>11/1/2006</i>	<i>2/1/2007</i>
1	V	14	Impact evaluation for first program year	1/1/2007	2/1/2007
1	V	15	Impact evaluation for second program year ^{***}	1/1/2007	1/1/2008
1	V	17	Annual update of energy impacts vs. goals	1/1/2007	3/1/2007
2	<i>II</i>	<i>3</i>	<i>Program, technology and sector priorities</i>	<i>2/1/2007</i>	<i>4/1/2007</i>
2	<i>III</i>	<i>4</i>	<i>Program & portfolio design</i>	<i>4/1/2007</i>	<i>6/1/2007</i>
1	V	19	Market effects studies (MT & Info programs)	7/1/2007	1/1/2008
1	I	2	Potentials analyses, baselines and updates ^{***}	8/1/2007	2/1/2008
2	<i>III</i>	<i>6</i>	<i>Portfolio review and approval</i>	<i>8/1/2007</i>	<i>10/1/2007</i>
1	V	20	Individual program future potentials	9/1/2007	12/1/2007
2	<i>IV</i>	<i>7</i>	<i>Final designs, program implementation and launch preparation</i>	<i>10/1/2007</i>	<i>1/1/2008</i>
2	<i>IV</i>	<i>8</i>	<i>Overall evaluation planning</i>	<i>10/1/2007</i>	<i>11/1/2007</i>
2	<i>IV</i>	<i>9</i>	<i>Regulatory review</i>	<i>11/1/2007</i>	<i>1/1/2008</i>
1	V	17	Annual update of energy impacts vs. goals	1/1/2008	3/1/2008
2	V		Program start-up	1/1/2008	7/1/2008
2	V	10	Evaluation contractor selection	1/1/2008	4/1/2008
2	V	11	Develop detailed program-level evaluation plans	1/1/2008	4/1/2008

Table 5.2: Continued

Track	Activity	Step	Description	Start Date	End Date
1	III	5	Program & portfolio selection	6/1/2008	8/1/2005
2	V	12	Process evaluations	7/1/2008	7/1/2009
2	V	18	Market effects baseline studies (MT & Info programs)	7/1/2008	11/1/2008
2	V	13	Measurement and verification	9/1/2008	7/1/2009
1	V	16	Impact evaluation for third program year	11/1/2008	3/1/2009
1	V	17	Annual update of energy impacts vs. goals	1/1/2009	3/1/2009
2	V	14	Impact evaluation for first program year	1/1/2009	2/1/2009
2	V	15	Impact evaluation for second program year ***	1/1/2009	1/1/2010
2	V	17	Annual update of energy impacts vs. goals	1/1/2009	3/1/2009
2	V	19	Market effects studies (MT & Info programs)	7/1/2009	1/1/2010
2	I	2	Potentials analyses, baselines and updates***	8/1/2009	2/1/2010
2	V	20	Individual program future potentials	9/1/2009	12/1/2009
2	V	17	Annual update of energy impacts vs. goals	1/1/2010	3/1/2010
2	III	5	Program & portfolio selection	6/1/2010	8/1/2007
2	V	16	Impact evaluation for third program year	11/1/2010	3/1/2011
2	V	17	Annual update of energy impacts vs. goals	1/1/2011	3/1/2011

*** Second year impact results feed step 2 in subsequent cycle.

The two-track process has the potential to inform consecutive or off-cycle program tracks by allowing for the use of evaluation results to refine the ex-ante or deemed savings estimates after the second and third year program evaluations are completed. It is possible to schedule and plan ex-ante savings updates to be coordinated with the second and third year program evaluation results that are needed to adjust these energy savings projections.

Likewise, the program-specific potentials analysis conducted within the program implementation and evaluation cycle can be used to inform the wider market potential assessments that are conducted periodically to assess the ability of the market to provide efficiency and demand resources. Using a coordinated approach to the implementation and evaluation of programs, in a sequenced series of coordinated efforts, means that the information needed to inform market potential assessments will be available at specific periods of time within each program cycle. It therefore becomes possible to plan and schedule potentials assessments to complement the information that is available to feed these assessments.

A Gantt chart showing the schedule for the two-track process is shown in Figure 5.5. Arrows showing interactions within and between the program tracks are shown in the Figure.

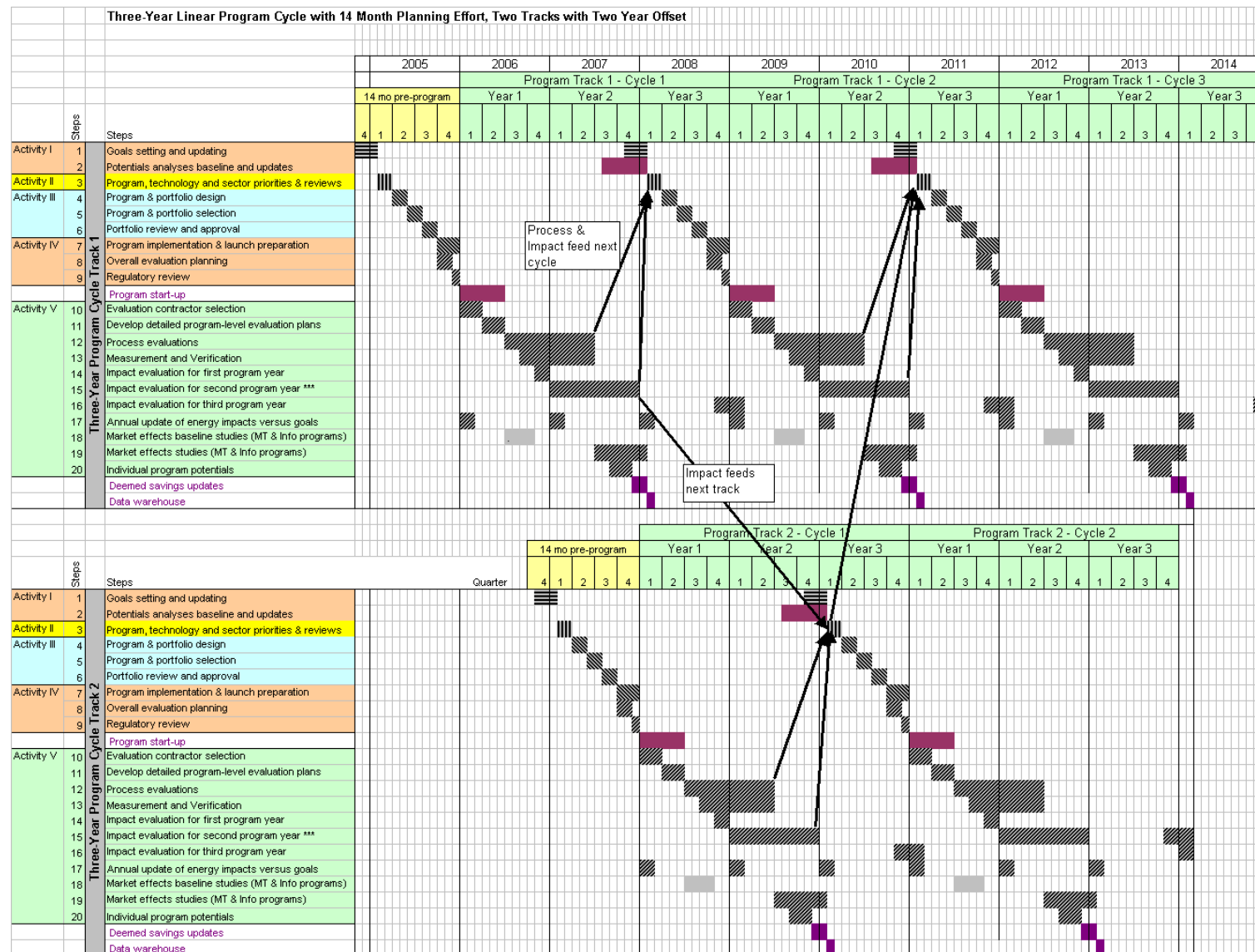


Figure 5.5: Example of Two Program Track Information Interactions

Chapter 6: Impact Evaluation

Preface

The Umbrella Roadmap (Chapter 5) provided some guidance on when an impact evaluation should be conducted. This chapter provides more details on how to conduct impact evaluations and identifies the studies and general steps needed to perform an impact evaluation. As such, this chapter provides a useful review of the impact evaluation methods and issues for policy makers and implementers.

California has a long history of conducting impact evaluations, dating back to the 1980s. The California Demand-Side Management Measurement Advisory Committee's (CADMAC) protocols, first written in 1993 and amended several times through 1998, provided the framework for conducting evaluation studies of programs funded by the state IOUs and filed with the CPUC. The purpose of these evaluation studies was to verify the energy and demand savings provided by these programs as part of a performance earnings process. The evaluation studies and methods prescribed by the CADMAC protocols were impact evaluations. Other types of evaluation, such as process and market evaluations, were not addressed by the Protocols.

Impact evaluation was the focus of California's Measurement and Evaluation Protocols and will continue to be a cornerstone of evaluation in California. Impact evaluations focus on estimating the "net" energy and demand savings of a program and its cost-effectiveness, thereby supporting California's program design, planning, and evaluation objectives. However, while the basic impact evaluation objectives have not changed, the operating environment has changed significantly in recent years and will affect the practice of impact evaluation:

- **Energy price volatility:** the temporal variation in energy prices has made impact evaluation requirements potentially more stringent. Simple reports of annual total energy (kWh) and peak demand (kW) savings may not be sufficient for program managers and policy makers who are interested in impacts at the hourly or daily level.
- **Energy price uncertainty:** the uncertainty in the balance of future supplies and demand and the effects of known and unknown market pressures make projections of future program benefits more difficult. Evaluations that use new pricing and price forecasting methods, with their own time/price assumptions, may be necessary.
- **Introduction of non-utility implementers:** the audience for impact evaluation has shifted from highly skilled evaluation professionals at the CPUC and IOUs to include a diverse group of non-utility program implementers with varying backgrounds in evaluation. For many of these new entries, evaluation is a new and confusing field, so that impact evaluations will need to be carefully described.

- **Appropriate evaluation scope for program:** Many non-utility programs deliver traditional energy savings technologies in innovative ways. These programs often have fairly small budgets, and do not have the program budgets for extensive impact evaluation requirements. If they are delivering new technologies or new configurations or new uses of existing technologies, impact evaluations for these programs may be important.
- **Relevant research questions:** New program design and delivery mechanisms will require a new set of research questions across each type of impact evaluation.

As the environment has changed, the following issues will be very important for impact evaluation:

- **Market noise:** The impacts of a particular program in a market filled with many implementers offering similar programs under different names with different incentive structures and marketing methods make estimating the influence of any particular program problematic. Identification of non-participants may be difficult, since customers may not be able to discern between the various programs operating in the marketplace, and may not accurately recall what influences various programs may have had on their decision processes or even remember the program in which they participated.
- **Spillover effects:** Market transformation efforts and the market impacts of the statewide IOU programs highlight the need for developing new methods for estimating spillover effects into the non-participant population (“free drivers”). The level of free drivers may be extensive. Enhanced marketing of energy efficiency programs by non-utility implementers further increases the technical challenges of measuring spillover.
- **Free ridership:** Free riders are project participants who would have installed the same energy efficiency measures if there had been no program. How free ridership is handled is a critical component of making the evaluations cost-effective and accurate. Uncertainty surrounding free ridership is a significant component of net energy and demand savings uncertainty; decreasing this uncertainty through additional studies could be expensive.
- **Accuracy requirements:** The accuracy requirements for impact evaluations may need to be relaxed under certain conditions, depending on the ability to use statewide impact evaluation measure results, size of the effort, the relevant research questions, budget, and the decision processes tied to those questions.
- **Data source and analysis methods:** The suite of tools available to conduct impact evaluations will need to evolve as the research questions change. Traditional engineering analysis, building energy simulations, metering and monitoring protocols, and statistical analysis of billing and survey data may need to be combined in new ways to answer emerging research questions.⁴⁰ The application of these tools will need to be described in a direct and straightforward manner to meet the needs of a diverse group of program planners and implementers. In some cases, new research questions may call for new analysis tools. Data

⁴⁰ See (Vine and Sathaye 2000).

- standardization efforts should be used whenever possible to gain the maximum usefulness of any data collection effort.
- **Uncertainty:** Explicit approaches are needed to quantify uncertainty, rather than relying on single point estimates. See Chapter 12 on Uncertainty for more information.
 - **Integrated planning process:** Chapter 5, The Umbrella Framework, describes an integrated planning process that uses impact evaluation results to inform the planning process for the next round of programs. Conducting impact evaluations in this environment will place specific requirements on study timing and data reporting requirements.

Skills Required for Impact Evaluation

This chapter provides an overview of the methods used in billing and engineering analysis. The primary distinctions between the billing analysis path and the engineering analysis path are that the former uses both pre- and post-installation billing data and relies more on statistical/econometric methods. The billing analysis section provides background and references for policy makers, evaluation managers, and evaluators new to this field but with some knowledge of regression analysis. It is not meant to provide a “how-to” manual. The discussion here is meant instead to provide an overview and a perspective on how regression analysis methods are applied in the task of evaluating energy efficiency programs.

The authors refer readers to formal regression analysis training and textbooks in statistics and econometrics to ensure proper use of the methods discussed. This includes examining alternative correction methods for violations of regression assumptions, and a more thorough understanding for the general application of these methods.⁴¹ The reliance on statistical/econometric methods requires personnel trained in these methods, as does the critical review of evaluations using regression methods.

The engineering analysis section provides background and references for policy makers, evaluation managers, and evaluators new to this field, but with some knowledge of basic energy engineering principles. References to the literature are provided for more information on engineering methods. The engineering analysis section covers both simple engineering methods and building energy simulation models. Simple engineering equations can be understood and used by most people with a general science

⁴¹ Statistics textbooks that include significant discussions of regression analysis and econometric textbooks can be found in almost all college bookstores and larger libraries. Most of these probably provide adequate treatment for the use of regression in billing analysis. Two examples include: *Econometric Models and Economic Forecasts*, (Pindyck and Rubinfeld 1981); and *Elements of Econometrics*. (Kmenta 1971).

background.⁴² Building energy simulation models generally require personnel with a graduate degree in mechanical or architectural engineering.

Introduction and Key Issues

Impact evaluation estimates the amount of electrical energy, electrical demand, and fuel energy saved due to a program, thereby verifying whether the expected program energy and demand savings are actually occurring in the field. Impact evaluation, therefore, involves estimating a **change** in energy use. Since it is possible to only directly measure consumption, to estimate savings one must observe the energy use characteristics of a program participant over time and from this generally infer what the energy consumption of the participant might have been in the absence of the program.⁴³

The primary purpose of impact evaluation is to obtain the most accurate and unbiased estimate of energy and demand savings due to the program. Methods used in impact evaluation can be useful for many other purposes in planning, load research, market research, and as information to work with process or market evaluation results. However, it is important to ensure that other desires do not lead the evaluation away from obtaining the most accurate and unbiased estimates achievable within a reasonable cost for the evaluation needs and uses.

There are two types of savings estimates that are normally desired from impact evaluation: gross savings and net savings. **Gross savings** are calculated for program participants relative to their prior participation usage. **Net savings** controls for savings that would have occurred for these participants over the same time period whether the program was offered or not. Estimating net savings generally requires the use of a comparison group as a proxy for what the participants would have done absent the program, or self-reported information on what would have happened in the absence of the program when comparison groups cannot be reasonably identified.

⁴² The Association of Energy Engineers (AEE) offers a certificate for a Certified Energy Manager (CEM). The material covered in the CEM program is good background for understanding engineering methods in evaluation.

⁴³ Many of the basic references on impact evaluation were written for EPRI in the 1990s. These references are not easily available to the general public. However, there are several references from the International Energy Agency <<http://dsm.iea.org>> that cover general evaluation topics that are in the public domain. See *A European Ex-Post Evaluation Guidebook for DSM and EE Service Programmes*, (AIS 2001); and *Evaluation, Verification, and Performance Measurement of Energy Efficiency Programmes* (* Violette 1995) for more information. The international climate change mitigation effort has also produced some excellent general references. See *Guidelines for the Monitoring, Evaluation, Reporting Verification and Certification of Energy Efficiency Projects for Climate Change Mitigation*. (Vine and Sathaye 1999). *Handbook of Evaluation of Utility DSM Programs* is an early public-domain reference that provides good background material. (* Hirst and Reed 1991).

Overall Impact Evaluation Objective

For the purposes of this Framework, the overall objective of impact evaluation is to estimate the net change in electricity and gas consumption and electricity demand on a program level. Electricity consumption impacts are generally defined in terms of kWh saved over some specified period. Fuel consumption impacts are generally defined in terms of therms of natural gas, gallons of propane, gallons of fuel oil, and so on. Electrical demand impacts are generally defined in terms of kW savings averaged over a short interval (e.g. 15 – 60 minutes). Demand impacts may be specified for a particular day and hour, or may be reported as the maximum savings occurring over a specified time period (e.g. summer on-peak hours).

Impact evaluation results are generally reported by costing period, which break the year into several bins depending on the season of the year and the utility system loading during these seasons. An example of the costing period definition used by Pacific Gas and Electric Company for commercial customers is shown in Table 6.1.⁴⁴

Table 6.1: PG&E Commercial Customer Costing Period Definition

Costing period	Summer (May 1 - Oct 31)	Winter (All Other Months)
Peak	12 PM to 6 PM Weekdays	none
Partial-Peak	8:30 AM – 12 PM Weekdays 6 PM – 9:30 PM Weekdays	8:30 AM – 9:30 PM Weekdays
Off-Peak	9:30 PM – 8:30 AM Weekdays All Weekends and Holidays	9:30 PM to 8:30 AM Weekdays All Weekends and Holidays

The costing periods generally vary by utility and customer class, which makes it difficult to define a single costing period for evaluation results reporting. The costing period definitions used by utility planners may also change over time depending on price and load volatility. Calculating load impacts on an 8760 hourly basis provides the flexibility to apply the results to a variety of costing period definitions.⁴⁵

Coincident peak demand impacts are generally evaluated at the specific day and hour of maximum system demand. For non-weather-dependent measures, the daily peak load generally corresponds to a summer workday at 4 pm. For weather-dependent measures, it generally refers to 4 pm on the hottest summer workday.⁴⁶ The precise calendar day associated with the peak outdoor temperatures varies by location, and not all locations are expected to have the maximum temperature occur on the same day. Building energy simulations are generally driven with typical or long-term average weather data rather

⁴⁴ This definition is used by PG&E in their commercial time-of-use rates. It varies somewhat from the definition of on-peak savings from the CPUC Energy Efficiency Policy Manual (noon – 7 pm, M-F, June-September) (* CPUC 2003).

⁴⁵ A recent CPUC study on avoided costs (Energy and Environmental Economics, Inc. 2004) recommended a Time dependent valuation (TDV) methodology for estimating avoided costs on an 8760 hourly basis. This methodology places a value on energy savings that varies by hour of the year and location to “better reflect the true avoidable costs to users, the utility system and to society.”

⁴⁶ This definition is consistent with the *California Statewide Commercial Sector Energy Efficiency Potential Study* (Coito and Rufo 2002).

than actual weather data for a particular time period.⁴⁷ The California Energy Commission (CEC) Climate Thermal Zone (CTZ) weather data sets are commonly used long-term average weather data set. An analysis of 8760 hourly weather data records in the CTZ files indicated the following days corresponding to peak temperature events in each of the sixteen climate zones.⁴⁸

**Table 6.2: Day of Year for Coincident Peak Analysis
(Based on 1995 Calendar year)**

CTZ	Peak Day
1	7/21/95
2	7/24/95
3	7/18/95
4	7/18/95
5	9/5/95
6	9/8/95
7	7/31/95
8	7/20/95
9	8/8/95
10	8/14/95
11	8/3/95
12	7/24/95
13	8/15/95
14	8/7/95
15	7/21/95
16	8/7/95

Demand impacts can be estimated from an analysis of billing data when the data include actual building peak demand for each costing period, although the coincident peak is generally not reported. Metering studies (e.g. pre/post time-series measurements of electrical demand)⁴⁹ can be used to estimate demand in sites with permanent or temporary recording interval demand meters. Demand impacts can also be estimated from energy impacts by applying a series of standard load shapes to allocate energy consumption into costing period bins. Sources for the load shape data include energy savings load shapes by measures that will be included in the California Database for Energy Efficiency Resources (DEER),⁵⁰ or from standard measure energy consumption load shapes applied to energy consumption savings estimates.⁵¹ Note: Load shapes based on building energy

⁴⁷ An important exception is when simulation models are calibrated to actual billing or end use energy consumption data. See Chapter 7 – Measurement and Verification for more information on simulation model calibration.

⁴⁸ Analysis conducted by AEC for PG&E CASE initiative study.

⁴⁹ The role of interval demand metering in measuring energy and demand impacts is covered in more detail in Chapter 7 – Measurement and Verification.

⁵⁰ (Xenergy et al. 2001).

⁵¹ Demand impacts in the latter case have long been estimated at PG&E by using a multiplier on the energy consumption savings that PG&E calls the “H” factor. H-factors vary by end use, building type, and climate zone.

consumption may not provide a good estimate of the load shape of measure savings when energy savings are not linearly proportional to consumption.⁵²

Units of Measure

Energy metrics for impact evaluation are defined as annual end use energy and demand savings by costing period. The CADMAC Protocols required that impact data be normalized according to designated units of measure. The designated unit of measure specified by the Protocols varies by the program type and end use. Typical units include impact per dwelling unit (for residential programs), impact per lamp (for lighting programs), and impact per unit floor area for commercial programs. These designated units of measure were selected to help load forecasters at the California Energy Commission predict the impacts of demand-side management programs on the long-term statewide electrical load forecast.

Conducting impact evaluations within the context of an integrated planning process as described in Chapter 5 will require results reporting in a format compatible with the activities included in the process. Efficiency potential studies and program planning activities generally require impacts on an individual measure or measure bundle level. Portfolio analysis generally requires impacts on a program level. Load forecasting occurs at a statewide and utility level, and the data needs of the load forecasting models used by utilities and CEC planners should be considered. Measure and building specific information reported during measurement and verification (M&V) can provide the detailed engineering data useful to inform the efficiency potential and program planning processes.

If the impact evaluation and its related M&V efforts are designed to provide technology-specific energy savings estimates, then the evaluation should include an effort to coordinate the results with updates to the DEER. The evaluation plan should address the technology-specific DEER data that can be updated by the evaluation and provide information on how the reliability of the data in the DEER is improved as a result of the evaluation. Results used to update the DEER can come from M&V studies, impact studies or both. Unit savings estimates consistent with the DEER schema or other engineering data supporting the refinement of the DEER unit energy savings are desired. If the evaluation results should not be used to update the DEER, the evaluation plan and the evaluation report should indicate why the data should not be used to update the DEER. See Chapter 7 for more information on measure-level reporting.

Interactions Between Impact Evaluation and Planning Processes

Impact evaluation is viewed as one part of a continuous process of program planning, implementation, and evaluation. Thus, the results of impact evaluation studies do not stand alone, but are used as inputs into planning and improving future programs. See

⁵² A simple example is the economizer, or “free cooling” cycle of an HVAC system. Energy savings from these systems generally occur during periods of reduced HVAC loads.

Chapter 5, Umbrella Framework, for a description of the integrated program planning and evaluation process.

Selecting Methods

Chapter 5 provides information for how to decide when an impact evaluation is appropriate. This decision process is an important part of evaluation planning. Chapter 5 provides a broader examination of evaluation planning needs and the overall evaluation planning process.

The next question is what impact evaluation methodology or methodologies to use. This decision will be based on several factors. Some of these may include:

- Type of program
- Data collection costs and available budget
- Size of program
- Program evaluation history (verify/triangulate with another method, questions posed from past evaluations, etc.)
- Demonstrated effectiveness of methods when applied to other similar programs
- Whether the evaluation is of a single program or a consolidated evaluation of several programs

Billing analysis will tend to be preferred when:

- Both pre and post-retrofit billing data are available
- Expected program impacts can be expected to be observed in a billing analysis (e.g., at least 10% of total consumption, depending upon method used, cleanliness of billing data, and accuracy of measured variables in analysis)
- The analysis is of a program with larger numbers of participants that are more homogenous

Engineering analysis will tend to be preferred when:

- No pre-measure billing data is available, e.g., new construction
- Expected impacts are too small to likely be observed in a billing analysis (e.g., less than 10% of total consumption)
- The programs has a small number of participants or unique measures, e.g., with industrial process improvements
- The programs has significant investments in engineering methods within the program that can provide cost savings for a similar evaluation, e.g., programs that include substantial engineering M&V or building energy simulation modeling

Larger programs may want to use a combination of engineering and statistical methods, given the size of the resources being generated. It might also be useful to have programs, where possible, alternate evaluation methods during each evaluation cycle. Alternating

methods can help provide greater perceived confidence in the results when these results are similar.⁵³

Appendix C provides a sample set of evaluation planning guidelines that examine various program typing categories and how these can be used to set guidelines for selecting the impact methodology. It then provides a table where these criteria show the types of impact, M&V, and net-to-gross method that would then be selected.

A very general presentation of this initial decision process is graphically displayed in Figure 6.1.

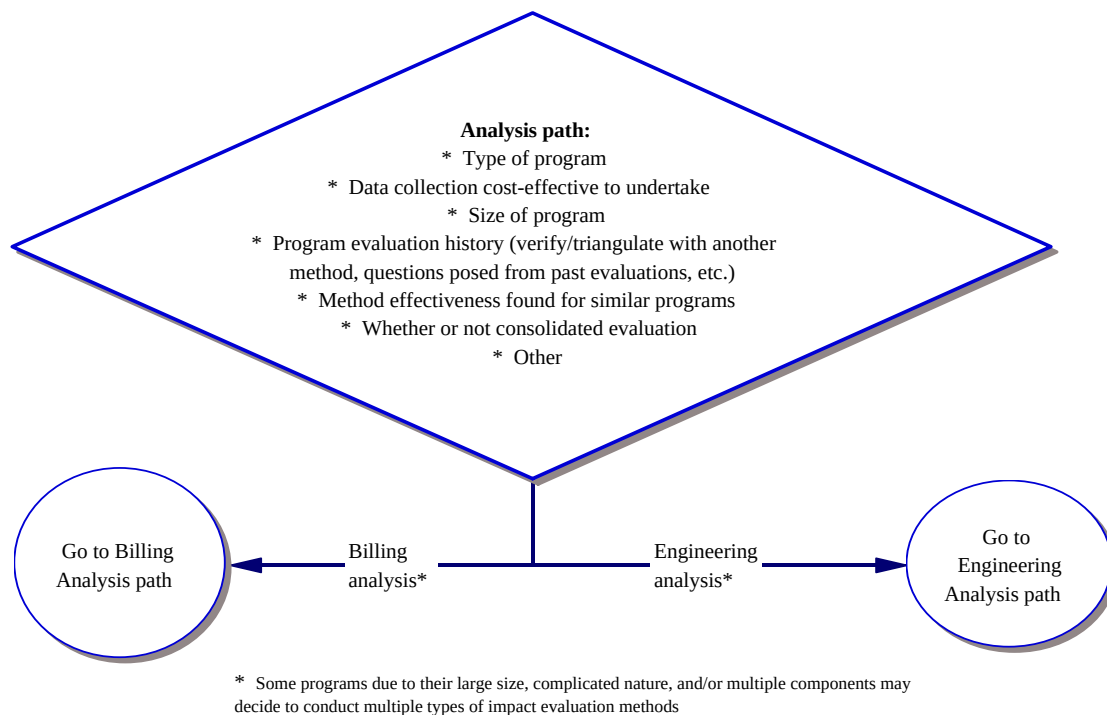


Figure 6.1: Example of General Decision Process for Selecting Impact Evaluation Approach

The Billing Analysis Path

This section of the Impact chapter provides a discussion of the critical issues that need to be assessed and addressed, wherever possible, in order for billing analysis results to be

⁵³ The calculation of statistical confidence is not affected. However, as noted in Chapter 12 on Uncertainty, bias is of significant concern and is immeasurable. Obtaining similar estimates from alternative methods can help provide evidence of a lack of bias if the potential sources of bias are not correlated between the two methods used.

used for impact evaluation. The section provides numerous examples of energy efficiency program evaluations applying these different methods, regression diagnostic tests, and corrections. There are many more possible examples available. Those presented here aim to provide a starting point for evaluators and reviewers, to use alongside regression textbooks and research, when more information is needed for a specific analysis issue.

The authors are not recommending one particular type of analysis method or regression model for billing analysis. As was done in the prior Measurement & Evaluation Protocols,⁵⁴ the Framework provides an overview of the reasonable methods that have been used, and more importantly, points to the need for evaluation professionals to use critical thinking to assess measurement issues and model assumptions, in order to provide the highest quality, most reliable impact estimates at a reasonable cost, given the nature of the program, data availability, and the evaluation questions to be addressed.

The Billing Analysis Roadmap is presented in Figure 6.2. It provides an overview of the likely methods and decision steps to be made. Nonetheless, this is a somewhat simplistic view of an analysis process that includes looking forward to anticipate data needs for data collection, and method and estimation adjustments as needed based upon early examinations of the data at hand.

The various billing analysis methods, literature references, and examples are presented in this section. The net-to-gross analysis methods, issues, and references are provided in a later section near the end of this chapter.

⁵⁴ *Protocols and Procedures for the Verification of Costs, Benefits, and Shareholder Earnings from Demand-Side Management Programs* (* CADMAC 1999) referenced as the Measurement and Evaluation Protocols.

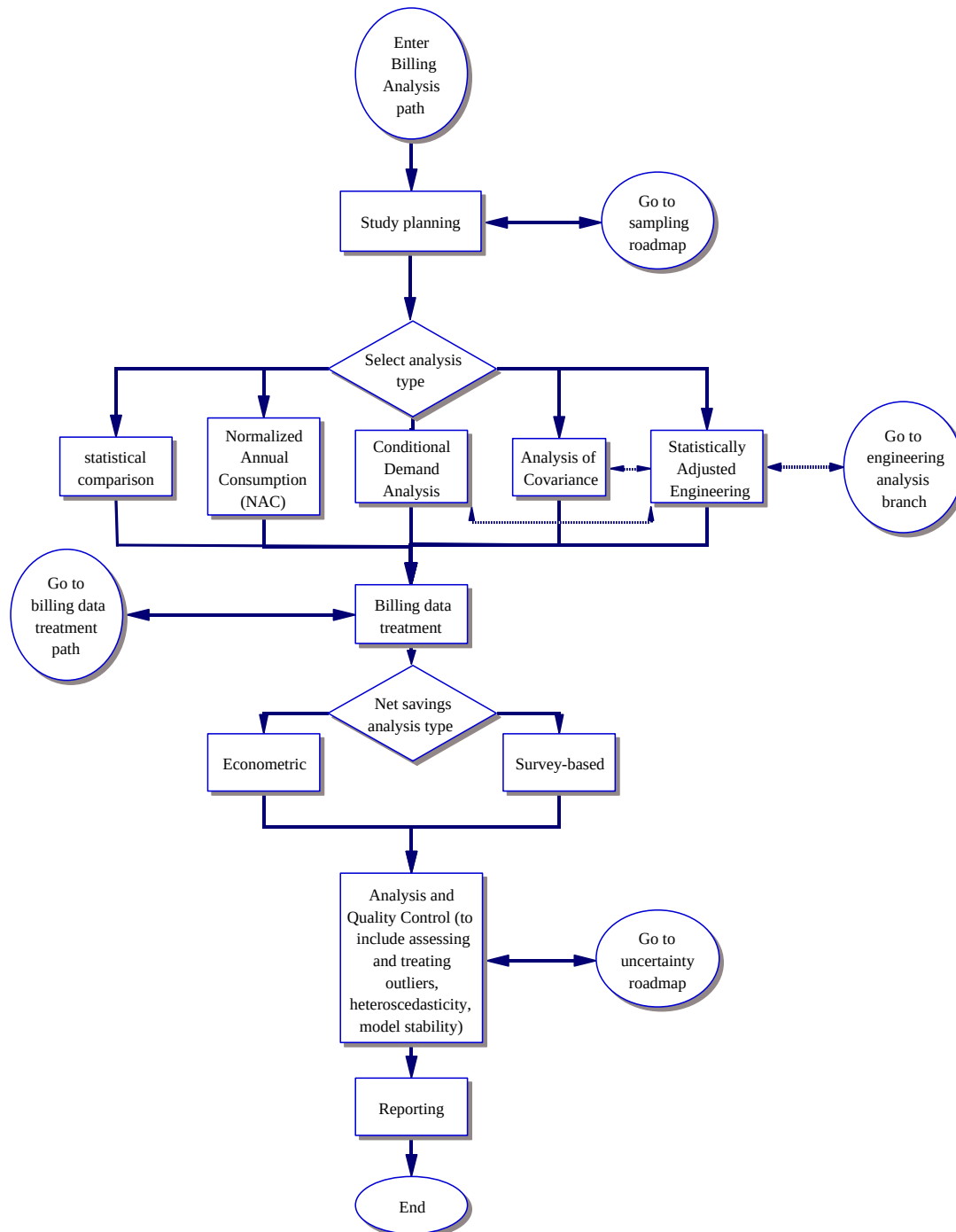


Figure 6.2: Billing Analysis Roadmap

Simple Aggregate Pre-Post Comparisons

The simplest use of billing data is to estimate program impacts by comparing average usage from the pre-installation billing data to post-installation bills. This is the One-Group Pretest-Posttest Design as described in the classic research design text by

Campbell and Stanley.⁵⁵ As discussed in this text, this simple research design contains many threats to the validity of its results.

One of the larger issues in using billing data (after the billing data have been cleaned) is that one of the largest drivers of energy consumption is the weather, and weather can vary significantly from one time period to another. This makes a pre-post comparison without adjusting for effects of weather generally meaningless. Accordingly, billing analysis normally adjusts for the weather in the pre- and post-installation periods.

A shortcoming of aggregating a customer's monthly usage to annual usage for pre-post analyses is that a lot of information is being "thrown away" that could be used in the analysis.⁵⁶ Having many monthly bills in the pre- and post-installation periods can provide more information for conducting the evaluation analysis, adding to its depth and, potentially, its reliability.

Another improvement on the pre-post design is the use of a matched comparison group in lieu of a simple comparison group.⁵⁷ A well-matched comparison group can be used to estimate what would have occurred in the absence of the program intervention (the counter-factual). Such a comparison group can come closest to serving as a control group, although the validity threat remains as some unmeasured factors relevant to measurement of effect might have been missed. A well-matched comparison group over the exact same time period might even mean that changes in the effects of weather, overall economy, and other external factors are already taken into account.

Comparing Pre-Post Billing Data for Programs with Experimental Design

Most of the regression discussion for billing analysis revolves around trying to isolate the program impacts from other elements that drive changes in energy use. With a true randomized experimental design there would be a control group rather than worrying and testing how well the comparison group matches the participant group. It would also provide net savings, avoiding the self-selection bias issues that dominate the discussion in the net-to-gross methods section near the end of this chapter.

Some have argued that publicly offered programs can not restrict who participates, that it may become a political problem and legal liability. It may be possible to create random experimental designs (or quasi-experimental design) within program design without creating political problems. The participation application process and program design offer opportunities for using experimental design that minimize some potential pitfalls in impact evaluation, such as the following:

⁵⁵ *Experimental and Quasi-Experimental Designs for Research*. (Campbell and Stanley 1963).

⁵⁶ *An Evaluation of Statistical and Engineering Models for Estimating Gross Energy Impacts*. (* Ridge et al. 1994).

⁵⁷ In a true experiment, a control group is established through random assignment between the treatment group and the control group. Seldom is this possible in social science interventions. A comparison group may be used as a "control group." Using the term comparison group helps to remind users that assessing how close the comparison group mimics the treatment group on all relevant variables to the treatment or selectivity of group selection is important.

- When people apply to participate, they are required to agree to participate in the study associated with the program whether they are selected to participate in the program or not. Then a random sample of these qualified applicants (those agreeing to the study) are selected to participate. Thus, the comparison group would be the same type of people as the participant group, and the program effects would be relatively simple to calculate. This is an especially defensible method if the number of potential participants is greater than resources can support.
- A program could accept all prospective participants, but assign them to a random participation time period (which could be multiple periods to minimize effects from month or season). The “not yet participants” form the comparison group, while the “already participants” form the treatment group. Billing analysis across these groups would not have the self-selection issues discussed in the net-to-gross section below.
- Matched community lists could be used in a neighborhood program. Then a random sample of communities could be developed to receive the program in year one, another in year two, and a third in year three as a phased roll-out. Effects due to the different time periods would still need to be examined. However, there are still significant advantages in using this method, for evaluation and potentially for program implementation, over a shotgun approach where the effort is a flat promotion across all communities rather than targeted efforts.
- An informational billing insert could be sent to a random sample of customers. Then surveys for recipients and non-recipients could be conducted for actions taken, or billing comparisons could be made.
- In-store rebates or in-store informational displays could be shown on random days, and sales of targeted materials affected by those days, relative to other days, could be calculated.
- An advertising rollout could be randomized across market areas. Then program/action response by market area and how that response changes with the advertising level of effort and/or message could be assessed.

Most likely, the program designs described above would initially cost more. At the same time, evaluation costs might be significantly reduced while providing potentially more defensible results. There is very little experience with experimentally designed energy efficiency programs. The advantages suggest, nevertheless, that a combined approach between program design and evaluation would be worth pursuing.

Types of Billing Analysis Models

As noted earlier, energy use analysis needs to adjust for the impact of weather on energy use. Weather adjustments can occur within a normal billing analysis framework by including weather variables, such as heating and cooling degree days. Weather normalized energy consumption can be estimated through regression for this type of comparison and is referred to as normalized annual energy consumption (NAC).

There are a variety of ways to adjust for weather including billing analysis models that perform individual regression analysis for each participant to adjust for weather and then provide the average program impacts from this grouping of regressions. One of the first and most popular of these methods in the late 1980s and early 1990s was the Princeton Scorekeeping Method (referred to as PRISM).⁵⁸ Given the popularity of this particular model, other methods that perform in the same manner are often called PRISM-like models.

Conditional demand analysis (CDA) describes a type of billing analysis in which observed energy consumption is estimated as a function of major end uses. A simplified model is presented here with binary (dummy) variables representing the presence or absence of end uses.⁵⁹ The resulting coefficients represent the marginal contribution to overall energy use associated with each end use. The dependent variable in these models is energy consumption (often monthly consumption).

The regression equation for a CDA model would appear as:

$$E_{it} = B_0 + B_1Ht_i + B_2AC_i + B_3WH_i + B_3CD_i + \sum_{K=1}^K B_{Kn}K_n + e_i$$

where:

- E_{it} = Energy consumption for customer “i,” in month “t.”
- B_0 = Constant picking up energy consumed through unspecified equipment.
- Ht_i = Dummy variable = 1 if customer “i” has electric space heating and 0 if not.
- AC_i = Dummy variable = 1 if customer “i” has air conditioning and 0 if not.
- WH_i = Dummy variable = 1 if customer “i” has electric water heating and 0 if not.
- CD_i = Dummy variable = 1 if customer “i” has electric clothes dryer and 0 if not.
- K_i = Dummy variable = 1 if customer “i” for each electric appliance in the model (and 0 if not), for appliances 1 to n.
- e_i = Statistical error term, for unexplained variance in observed energy consumption.

This type of regression model was developed to predict energy use and explain energy use to customers. CDA-type models have often been used to estimate unit energy consumption (UEC). The model above, however, has frequently been found to be too simplistic. UEC’s are often not linear and may depend upon a host of variables in complicated ways. A second problem lies in the nature of incorporating many variables,

⁵⁸ There are several papers in *Energy and Buildings* dedicated to this approach. Some of these include: “Exploratory Scorekeeping for Oil-Heated Houses,” (Fels et al. 1986); “A Midwest Low-income Weatherization Program Seen through PRISM,” (Goldberg 1986); and “The Applicability of PRISM to Electric Heating and Cooling.” (Stram and Fels 1986).

⁵⁹ “The Total and Appliance-Specific Conditional Demand for Electricity in the Household Sector.” (Parti and Parti 1980).

one for each electric using device. The possibility of over-specification and multicollinearity then becomes a real issue. In fact, a simplistic representation of “dumping” many dummy variables in for every appliance has at times been found to produce impossible estimates such as negative UECs. These problems are especially troubling where the primary evaluation question is not the predictive ability of the model for energy use but accurate and reliable estimates of energy savings or UECs.⁶⁰ Yet, careful work with CDAs that test and correct for these issues can make it a useful tool in the evaluation toolkit.

A study supported by San Diego Gas & Electric demonstrated through Monte Carlo simulation studies that a more direct estimation of savings than CDA could be as accurate in estimating savings as a CDA model with various CDA-type variables. Yet, it also demonstrated that if there is either measurement error for the variables added to the CDA model (errors-in-variables bias) or misspecification of the model then the more direct estimation of savings method is significantly more accurate.⁶¹

The term CDA has occasionally been expanded to cover a wide range of hybrid models that incorporate program data, consumption data, and engineering data. It is important to limit application of the term CDA to its original, more precise meaning. In this way, impact evaluation reporting can more easily be explicit about the type of model used in billing analysis.

In the traditional interpretation of terminology, a “Change Model” is different from a CDA Model. The CDA Model is designed to explain energy uses, while a Change Model is designed to explain changes in energy usage. That means there is a difference (or change over time) being explained by the model. This can take the form of having the change in energy consumption (pre versus post) as the dependent variable (e.g., December pre-retrofit usage – December post-retrofit usage), or having consumption as the dependent variable and pre-retrofit consumption as one of the independent variables.

⁶⁰ An OLS regression procedure uses only the variation unique to a regressor as it relates to the dependent variable. This means that if two or more regressors are highly correlated for the purpose of calculating the coefficients, the common variation is ignored. The overall variation is included in the R^2 and the overall estimate of usage may still be predictive. However, the coefficient of the regressors has high variance and may produce coefficients that are not reflective of the true effect. Given the coefficient of the regressors are the actual evaluation goal in either UEC estimation or energy savings estimation (from either a participation dummy variable or SAE variable), this is completely counter-productive. See the discussion below on misspecification and multicollinearity. Also see standard regression/econometric texts for discussions in this area, e.g., *A Guide to Econometrics*, 3rd edition. (Kennedy 1996), pages 176 – 187.

⁶¹ The direct savings estimate is opposed to a CDA-approach which seeks to estimate baseline consumption and post-program consumption and using the difference as the estimate of savings. See *Appendix Z Simulation Study: Comparison of Alternative Methods for Measuring the Gross and Net Energy Impacts of Demand-Side Management Programs (with Addendum)*. (Schiffman and Engle 1993), available at www.calmac.org.

This type of model could be specified as:

$$E_{i, post} = B_0 + B_1 INSTALL_i + B_2 E_{i, pre} + B_3 W_{it} + \sum_{K=1}^K B_K X_K + e_i$$

where:

- $E_{i, post}$ = Energy consumption for customer “i” in post-periods
- B_0 = Constant picking up energy consumed through unspecified equipment.
- $INSTALL_i$ = Dummy variable = 1 for participant (installer) and 0 for non-participant.
- $E_{i, pre}$ = Energy consumption for customer “i” in pre-periods.
- W_{it} = Average weather for customer “i” in month “t,” as defined by that customer’s billing cycle.
- X_K = Vector of other explanatory variables.

The B_1 coefficient on the *INSTALL* dummy variable provides the measure of energy savings.

To control for differences between customers, other explanatory variables (the vector of X_K variables in the above) can be used in these models, such as variables that affect energy usage (e.g., square feet, operating hours, industry) or that affect energy usage over time (e.g., the price of energy, economic conditions).

A special type of explanatory variable, incorporating the engineering estimate of savings, provided a category of models called Statistically-Adjusted Engineering (SAE) Models.⁶² The regression coefficient in these models is the percentage of the engineering estimate of savings observed in changes in energy usage. For example, if the coefficient on the SAE term is 0.8, this means that the customers are on average seeing 80% of the savings realized from their engineering estimates. The SAE model can estimate realization rates for an overall customer savings estimate or for individual measure savings estimates if these estimates are placed within the model.

A basic SAE model would look as follows:

$$E_i = B_1 S_i + B_2 W_{it} + \sum_{K=1}^K B_K X_K + e_{it}$$

where:

- E_i = Average energy consumption for customer “i” from the billing data.
- S_i = The measure savings estimates in a SAE model for customer “i” in month “t” are used instead of the *INSTALL* variable seen in the model

⁶² “Statistically Adjusted Engineering (SAE) Models of End-Use Load Curves,” (Train et al. 1985); “The Economic Value of Energy-Saving Investments by Commercial and Industrial Firms.” (Train and Ignelzi 1987).

above. For a measure-level SAE model, as proposed here, a variable and coefficient B for each measure-level savings estimate would be used.

- W_{it} = Average weather for customer “ i ” in month “ t ,” as defined by that customer’s billing cycle.
 X_k = Vector of other explanatory variables.
 $B_1...B_n$ = Estimate coefficients.
 e_{it} = Statistical error term, for unexplained variance in observed average energy consumption.

This technique was originally intended for use in billing analysis to calculate gross energy savings. It provides a method to incorporate all available information (most importantly engineering estimates) into the regression, while providing feedback on the accuracy of the engineering estimate. Confusion has occurred, as some have tried to re-interpret the SAE approach for analyzing net savings (e.g., with comparison group data in the regression), and others have said that the SAE was a simplified form of an expanded CDA Model. The authors recommend that evaluators use a “back to the basics” approach when using terminology to describe the regression methods used. This should be less confusing to readers. (For example, a regression model was used, a CDA model, a CDA-SAE model, or a SAE Change model, etc.)

Most regression models are estimated as ordinary least squares (OLS), generalized least squares (GLS), or other forms of maximum likelihood estimation.⁶³ These methods generally produce similar results under similar circumstances. Generalized least squares, as its name implies, is a more generalized statistical equation. If the error term is normally distributed, both OLS and GLS may be identical to the maximum likelihood estimate (MLE).⁶⁴ There are differences in these estimation methods, however, that lead to the decision of which model specification is more appropriate for different circumstances. The more generalized the method, the more it can often be used to correct for different issues. At the same time, it can become more computationally difficult.

The billing analysis models discussed thus far can use OLS, GLS, or maximum likelihood estimation. Discussion, comparisons, and derivations for these regression models can be found in most graduate regression statistics or econometric textbooks.

OLS is the most commonly used of these techniques and the easiest to use in most statistical software packages. It is important that the estimator of the model’s coefficients is unbiased (centered on the correct answer) and consistent (that it approaches the exact

⁶³ In general, there are three methods of estimation: moments, least squares, and maximum likelihood. (Moments is not discussed in the Framework. This method uses the sample mean as an estimate of the population mean.) See *Elements of Econometrics*. (Kmenta 1971).

⁶⁴ The maximum likelihood estimate (MLE) is based upon the concept that different populations would generate a different sample. It is estimating what the population would look like that would most likely provide the sample values obtained. The technique provides estimators that maximize the probability of getting the sample, i.e., what estimators provide a population distribution that would maximize the probability of seeing the sample being analyzed. See pages 174 – 182 of *Elements of Econometrics* (Kmenta 1971), or the MLE discussion in any graduate econometrics textbook.

population coefficient as the sample size gets larger). According to the Gauss-Markov theorem, the best (minimum variance producing) linear unbiased estimator can be achieved with ordinary least squares, as long as four assumptions are correct. These are:

1. The relationship between the dependent and independent variables is linear;
2. The independent variables are fixed, non-stochastic variables (i.e., they are non-random variables known by the researcher);
3. The error term has zero expected value, and constant variance for all observations (i.e., the residuals fall uniformly around the correct answer, without a significant pattern being formed in the residuals according to unmeasured characteristics of the observations);
4. The error term is random and the errors (residuals) are uncorrelated to one another.

If these four assumptions are correct, OLS will provide a better estimator than GLS. But, if any of these assumptions are incorrect, GLS should be used. If OLS is inappropriately used in cases where one of these assumptions is violated, it will produce an estimator with greater variance than GLS would have produced. There are also common techniques that can be used along with GLS to correct for cases that violate these assumptions.

One billing analysis method that is designed to use GLS rather than OLS is Analysis of Covariance (ANCOVA) models. An ANCOVA model is also referred to as a “fixed effects” model. This model allows each individual to act as its own control. The unique effects of the stable, but unmeasured characteristics of each customer are their “fixed effects” from which this method takes its name. These fixed effects are held constant. The fixed effects nature of the model means the ANCOVA model does not need to include unchanging customer characteristics such as square footage, number of floors, equipment in the home, etc.^{65, 66}

Controlling for fixed effects controls the amount of variance (noise) the model is faced with, since each customer has a different baseload, a different response to weather, and a different pattern of consumption that changes over time. This approach also provides for a much closer fit to the data than most models as individual responsiveness is incorporated. At the same time, using individual responsiveness is more meaningful than including lagged usage variables. This method is similar to the PRISM-type models that

⁶⁵ *Loan and Whole House Rebates: Program Evaluations*, (Megdal et al. 1993b); and “Estimating Takeback (Comfort Increase) For a Low-Income Program, Loan Program, and a Single Family Rebate Program.” (Megdal et al. 1993a).

⁶⁶ “A Comparison of Model Specifications In a Billing Data Analysis of Impacts From a Commercial and Industrial Rebate Program,” (Sumi et al. 1993); “The Treatment of Outliers and Influential Observations in Regression-Based Impact Evaluation,” (Schutte and Violette 1994); “A Monte Carlo Based Comparison of Techniques for Measuring the Energy Impacts of Demand-Side Management Programs.” (Schiffman 1994).

allow separate baseline consumption for each customer and separate regressions for each customer.⁶⁷

The basic model framework for an ANCOVA SAE model is as follows:

$$E_{it} = B_1 S_{itj} + B_2 W_{it} + \sum_{K=1}^K B_{ki} + e_{it}$$

where:

- E_{it} = Average daily energy consumption for customer “i” in month “t,” from the billing data, with the consumption for the billing cycle, divided by the number of days in the billing cycle.
- S_{itj} = Dummy variable = 1 if customer “i” in month “t” had installed measure “j”; = 0, if the conservation measure had not yet been installed. For a SAE model, the measure savings estimates would be included in place of the “1” for the months after installation. For a measure-level SAE model, there is a variable for every engineering estimate “j.”
- W_{it} = Average weather for customer “i” in month “t,” as defined by that customer’s billing cycle.
- B_{ki} = For ANCOVA, customer “i,” included as own control for fixed-effects. The coefficient adjusts for the customer’s base usage as differentiated from the usage for the sector based upon the other variables in the model. This means there is a coefficient for every customer in the model, so a model with 500 customers would have 500 coefficients. Interacted with weather, the coefficient adjusts for the customer’s weather sensitive usage, as differentiated from the usage for the group as a whole, based upon the other variables in the model. In our example, this would add another 500 coefficients to the results.
- $B_1...B_n$ = Estimate coefficients.
- e_{it} = Statistical error term, for unexplained variance in observed average daily energy consumption, for customer “i” in month “t.”

The coefficient “B₁” will provide either the average daily consumption savings from the measures installation (standard billing analysis), or the percentage of the engineering estimate obtained for an SAE model - depending on whether a dummy variable is used, or whether all sample participants have program engineering estimates available for all measures installed. From this coefficient the average savings can be estimated.

Regression Summary

Most billing analysis conducted to evaluate energy efficiency programs is done with one of the above models. Variants include incorporating comparison groups in the analysis

⁶⁷ One of the differences between PRISM-like models and ANCOVA is the greater flexibility of the ANCOVA model, which allows separate customer-specific interactions with weather and other variables to be easily assessed. The GLS form generally used to perform ANCOVA in standard statistical software packages also allows corrections for various regression assumption violations to more easily be examined and controlled.

and using methods that attempt to control some of the selectivity bias that occurs in participation (i.e., not having the comparison group a perfect match for the participant group). There are also tests and methods used to correct for violations in the assumptions of the regression analysis.

In fact, most of the work that many econometricians do is intended to deal with the reality that the key assumptions of OLS are usually not met in field data. Most graduate textbooks on regression analysis spend the first quarter on the basics of regression and then the rest of the book is in providing ways to address violations in the OLS assumptions, model form, data, or other “problems” when using OLS.

One of the more critical assumptions in regression analysis is that the model specified is the correct model. Using regression analysis for causality assumes that the independent variables cause the actions being measured in the dependent variable, not just a correlation. It also assumes that this is the one and only true model for the cause of the movement seen in the dependent variable. This assumption is for the variables included, the mathematical form of the interaction, and the treatment of any non-random error term effects. Model specification is a critical component of regression analysis. Problems concerning model misspecification will therefore be examined again later in this section.

Accordingly, there is a need for experienced professionals to conduct billing analysis by critically analyzing their data and situation to be able to find when alternative models, tests, and corrections need to be employed. They must do so in a professional manner, reporting on their research design, logic, and potential issues (such as potential bias issues, as discussed in Chapter 12 on Uncertainty).

A few of the more common issues, tests, and corrections will be discussed below. The authors are not recommending any specific billing analysis model, specific tests that must occur, or methods for corrections when statistical issues are found. However, it is essential that the evaluator’s work be a credible professional job in line with the evaluation needs for that project and research results in the fields of regression analysis, econometrics, and energy efficiency evaluation.

Bayesian Analysis

The above discussion on billing analysis models has come from the more traditional statistical perspective. Another approach can be undertaken from a Bayesian perspective. (For a basic description of Bayesian statistics, see the sidebar provided in the Chapter on Uncertainty.) The foundation of the difference between the two paradigms is that classical statistics is based upon the notion of repeatedly drawn samples and the sampling distribution of the estimate for the true parameter, while Bayesian statistics is founded on subjective probability and the use of prior information.⁶⁸

Bayesian output can be quite meaningful in an evaluation framework, though the interpretation is somewhat different.

⁶⁸ While the SAE model uses prior information by incorporating the engineering estimates of savings, it is not conducted as a Bayesian analysis and has not been interpreted in a Bayesian manner.

“Instead of producing a point estimate of β , a Bayesian analysis produces as its prime piece of output a density function for β called the “posterior” density function. This density function relates to β , not B , so it most definitely is not a sampling distribution; it is interpreted as reflecting the odds the researcher would give when taking bets on the true value of β .”⁶⁹

Very little energy efficiency evaluation has occurred within a strict Bayesian approach.⁷⁰ This type of analysis may be applicable for impact evaluation, but the evaluation plan would need to show how the method would work for the evaluation being conducted, the advantages of doing so, and how it is to be conducted. Then it could be recommended to be part of the waiver process.

Critical Billing Analysis Issues and Quality Control Efforts

The perspective throughout this document has been to provide a discussion of the foundation that supports the roadmap and references for more information. It then directs evaluations to be conducted by skilled, professional evaluators for the tasks to be performed and to be done in a manner guided by the science in that area (e.g., engineering for engineering analysis, statistics for sampling, and regression analysis for billing analysis). An extremely prescriptive manual of methods would not allow evaluations to be designed that would best fit the program and evaluation circumstances. This is particularly true in the area of billing analysis.

In billing analysis, as in all other evaluation approaches, it is important to describe in the evaluation plans and reports the potential areas of bias, questions about the specification of models, sampling issues, etc. In a high quality impact evaluation, the evaluators ensure that the evaluation is conducted so that issues of specification, sampling, bias, etc., are specifically tested, analyzed, and addressed, either empirically or qualitatively (as is appropriate to the issue, program, and evaluation goals), and that they are disclosed for knowledgeable reviewers to understand.

The Measurement & Evaluation Protocols' Quality Assurance Guidelines presents a series of questions that needed to be answered as part of the evaluation analysis and reporting (pages 11-14). The Framework supports having these issues tested for, assessed, and addressed where needed for billing analysis. They include:

- Misspecification
- Non-random error term
- Non-random measurement error
- Autocorrelation

⁶⁹ *A Guide to Econometrics*, 3rd edition. (Kennedy 1996), page 201. For an introduction to Bayesian statistics see: *Bayesian Statistics for Evaluation Research: An Introduction*. (Pollard 1986).

⁷⁰ An initial assessment of the usefulness of using a Bayesian approach for impact evaluation is shown in “Impact Evaluation Accuracy and the Incorporation of Prior Information.” (Violette 1991).

- Heteroscedasticity
- Collinearity
- Tests for exogeneity and endogeneity
- Influential data
- Missing data
- Triangulation
- Weather effects
- Engineering priors
- Precision
- Comparison group

Model Misspecification

An easily overlooked regression problem is misspecification of the model. A critical assumption in regression analysis is that the model is correctly specified, i.e., that it represent the underlying process. In some sense, regression assumes that the regression model being tested is the one and only true representation of the process that determines the dependent variable. Using regression analysis for causality assumes that the independent variables cause the actions being measured in the dependent variable, not just a correlation. This assumption pertains to the variables, the mathematical form of the interaction, and the treatment of non-random error term effects. These are strong assumptions, and most practitioners realize that a regression model may be missing some variables and data (some of which may be important). It is important, however, not to get complacent about the imperfections, testing, and correcting for misspecification. Typical categories of specification errors include the following:⁷¹

1. Omission of a relevant explanatory variable.
2. Disregard of a qualitative change in one of the explanatory variables.
3. Inclusion of an irrelevant explanatory variable.
4. Incorrect mathematical form of the regression equation.
5. Incorrect specification of the way in which the disturbance enters the regression equation.

It is hard to know when a model is mis-specified. Specifying the model is generally based upon the relevant theory, experience, the literature on similar work, and testing alternative models. There are a variety of tests that can be utilized.⁷²

⁷¹ *Elements of Econometrics*. (Kmenta 1971), page 392.

⁷² *A Guide to Econometrics*, 3rd edition. (Kennedy 1996). There is also a 2003 edition that might prove useful with updating and additions from latest research.

Econometricians are trained in the theoretical underpinnings of the statistics and correct for a variety of non-random error term situations through statistical tests and corrections. An alternative approach to analysis has been recommended by Kennedy that it is important to consider. “Models whose residuals do not test as insignificantly different from white noise (random errors) should be initially viewed as containing a misspecification, not as needing a special estimation procedure.”⁷³

Misspecification can cause many problems. Although difficult to detect as a problem, model specification and alternative specifications should normally be assessed by the results they provide, the stability of the model, and the consequences in the regression diagnostic tests. It is important to recognize that energy savings estimates depend not on the predictive power of the model on energy use, but on the accuracy, stability, and precision of the coefficient that represents energy savings. The best evaluations show the alternative model specifications considered and how they performed.

Non-Random Error Term

The most common critical threats to the validity of the statistical estimate of an estimator discussed in literature are the cases where OLS is used while one or more of its assumptions are violated.⁷⁴ (These assumptions were stated in an earlier part of this section.) As noted previously, if OLS is inappropriately used in cases where one of these assumptions is violated, it will produce an estimator with greater variance than GLS would have produced. GLS and other techniques can be used to correct for cases that violate the OLS assumptions.⁷⁵

The ANCOVA model discussed above is one way to address the problem of the error term not being truly random. It does so through measuring the covariance among categorical variables. Often these types of models are divided into random effects models (or variance components models) and fixed effects models. Much of the work in this field involves providing the appropriate estimators for differing circumstances or assumptions about the components and relationships of the error terms. (See the cited articles for more detailed discussions of this type of work and its applications.)⁷⁶

As noted in the discussion above on billing analysis methods, efficiency billing analysis often involves cross-sectional time-series data. In other words, there are observations over time for an individual or firm (e.g., collecting utility bill data during pre- and post-installation periods), and there are data across individuals or firms. These types of

⁷³ *A Guide to Econometrics*, 3rd edition. (Kennedy 1996), page 77.

⁷⁴ Recognize, however, that this may be the more recognizable problem while misspecification could easily occur most of the time without being recognized.

⁷⁵ This is because the only assumed error structure in the generalized method is that the variance-covariance matrix of the error terms be multiplicative scalar and positive definite. *Econometric Models and Economic Forecasts*. (Pindyck and Rubinfeld 1981).

⁷⁶ (Balestra and Nerlove 1966; Maddala 1971a; Maddala 1971b; Hausmann 1978; Mundlak 1978; Hausmann and Taylor 1981; Lillard and Acton 1981; Aigner and Lillard 1984; Aigner and Hirschberg 1985; Jasso 1985; Amemiya and McCurdy 1986; Cornwell and Rupert 1988; England et al. 1988; Megdal et al. 1993a; Sumi et al. 1993; Schutte and Violette 1994; * Megdal et al. 1995b; Ozog et al. 1995b).

datasets have the possibility of containing error distribution issues originating from either the time dimension or the participant dimension.

Non-Random Measurement Error and Errors-in-Variables Bias

Each of the independent and dependent variables need to be appropriately measured to ensure proper results. If variables are measured with random error, the model has additional “noise” that may make ascertaining the correct model more difficult and lowers the precision achieved. If the variables are measured with nonrandom error that has any correlation with other variables in the model, then the regression coefficient can be biased. If an independent variable is measured with non-random error, then the results will be biased and inconsistent. If the dependent variable is measured with error, the variances of the estimates will be large. If it is measured with non-random error, then relationships being measured by the model are not those in the construct, and the results may be biased.

If the interest is the dependent variable, random measurement error in the independent variable is less important. The coefficient for the independent variable may be biased but the general error term picks up the random error in the measurement of the independent variable and the estimation of the dependent variable is still valid. Unfortunately, however, most billing analysis models are predicting consumption and the variable linked to the primary research interest, energy or demand savings, is often an independent variable. The measurement error in independent variables can bias their coefficients to zero and measurement error in multiple independent variables can provide bias in their coefficients of unknown direction and magnitude. This problem is often referred to as errors-in-variables bias.⁷⁷

Heteroscedasticity

Another important error assumption is the constant error variance or homoscedasticity. There are many cases where this assumption may not be true. To not have a constant error variance or unequal variances, where the error pattern varies systematically for different types of individuals or firms (potentially a missing variable problem), is heteroscedasticity. This is quite likely when some unmeasured characteristic limits the variability for one group versus another group. For example, this assumption is not likely to hold for commercial analysis where there are small firms and a small group of very large firms. Large firms/buildings have the likelihood of having greater variance (and error variance) in variables with a greater potential size (i.e., size is correlated with the variance). This is quite likely because there are many more items and people that can vary in their actions in a larger building/facility than a smaller one. This is a classic example often used in regression textbooks, even outside the efficiency evaluation field, as a common problem for cross-sectional studies.

⁷⁷ A simulation that provides evidence of the accuracy problems caused by the errors-in-variables bias for energy efficiency program evaluation can be found in *Appendix Z Simulation Study: Comparison of Alternative Methods for Measuring the Gross and Net Energy Impacts of Demand-Side Management Programs (with Addendum)*. (Schiffman and Engle 1993).

Given its common nature among cross-sectional studies involving facilities, it is recommended that billing analysis for commercial or industrial programs, at a minimum, test for heteroscedasticity. At the same time, there can be many different causes of heteroscedasticity (including income distribution) that would suggest that any high quality evaluation would at least examine this possibility. A relatively simple first test would be to plot and view the residuals against variables that are likely to be correlated with the error variance issues (e.g., facility size, income). There are also several formal tests available in most regression statistical software packages used that look for different relationships between the error terms and the independent variables, e.g., Goldfeld-Quandt (which looks at the magnitude issue in firm size), Breusch-Pagan (linear combination of known variables creating error disturbance), White test, etc.⁷⁸ Solutions are available for most problems. A common correction in GLS is to use weighted least squares, which is offered as an option in most regression statistical software packages.

Autocorrelation

Another common problem within this class of cross-sectional time-series models is autocorrelation.⁷⁹ Autocorrelation of errors is most common in time-series (due to the intrinsic relationship between the most recent prior period and the present measurement while unspecified variables are missing that would explain the underlying mechanisms for these changes). Autocorrelation, however, can also occur due to spatial autocorrelation (where firms located in the same area could have correlated errors due to an omitted variable for other factors influencing the firms that are correlated with their location).

The time-series nature of billing analysis makes it a prime candidate for problems with autocorrelation. There are, however, other logical ways in which billing analysis could generate autocorrelation problems. Let's suppose individuals living inland increase their air conditioner usage as temperature rises because either the heat is already in their buildings or due to their anticipation of how hot it may become inside the building. In contrast, individuals living closer to the coast are slower to respond to a temperature increase given an expectation of its temporary nature. If the weather variables cannot capture this complicated relationship and no geographic variable is used (and the model does not control for fixed effects, such as an ANCOVA model), then this circumstance could create a problem with autocorrelation given the non-random nature of the error term and its being correlated to geographic location.

Given that autocorrelation is relatively common in time-series models, a high quality evaluation would test for this. There are several tests and corrections available in common regression statistical software packages.

⁷⁸ *A Guide to Econometrics*, 3rd edition. (Kennedy 1996), page 118.

⁷⁹ Some econometricians believe that autocorrelation is a type of model misspecification.

Examples from Energy Efficiency Evaluation

The first example is a large commercial and industrial impact evaluation that used billing analysis as one of its evaluation methods.⁸⁰ Regression diagnostics were performed on all the models regardless of how “good” the initial modeling results appeared. The regression diagnostics used were based upon: testing the probability that the residuals were normally distributed, the skewness measurement, the kurtosis measurement, a Pearson’s correlation coefficient between the residual and the lagged residual, and an examination of residual plots against the predicted values, the savings estimate, average heating degree days, average cooling degree days, and time.⁸¹

The initial office sector energy model had significant problems with heteroscedasticity. The initial model had a probability of normally distributed residuals of 71 percent, a skewness measure of 2.43, and a residual plot showing one customer with consistently higher savings. This was solved by creating two models, a separate model for a large customer who had had much of its retrofitted space vacant in the post-period, and another model for the remainder of the sector. The correction of using two models allowed the sector model to have a probability of normally distributed residuals of 90 percent, and the skewness measure fell from over 2 to -0.8. These corrections found significant differences in the realization rates achieved for the savings estimates, proving the importance of this type of examination. One realization rate was created and used for the unusual customer and one for the rest of the office sector.

Within the same project, theoretical work was conducted on how trends in economic conditions would affect savings. This analysis showed that savings estimates would be overestimated (in terms of what otherwise would have occurred, or from the perspective of long-run expected realization rates) during recessionary trends. Savings estimates would be underestimated during strong periods of economic growth. The period being analyzed was known to be a strong recessionary period. This meant that, without considering this factor, savings would be overestimated.⁸²

Since the analysis was conducted during a recessionary “trend,” the evaluators hypothesized that the economy might be contributing to the “trend” problem seen in the autocorrelation tests. Autocorrelation was found; the residual and lagged residual had a Pearson’s correlation coefficient of 0.59. A trend variable was used to correct the autocorrelation and as a proxy for the decreasing consumption trend seen in manufacturing over the long analysis period.

⁸⁰ “The Importance of Using Analysis of Covariance (ANCOVA), Diagnostics, and Corrections within Billing Analysis for Large C&I Customers.” (* Megdal et al. 1995b), page 440.

⁸¹ Skewness is the measure of the departure from symmetry, with the median being different than the mean and skewness being defined whether the tail is elongated to the left or to the right. Kurtosis is the extent to which the sample is peaked (leptokurtic) or flatter (platykurtic) than a normal distribution. Pearson’s correlation coefficient is one of the most common methods of assessing the nature of the relationship between two variables.

⁸² “The Changing Economy as Part of DSM Impact Evaluations: Evidence from a Large C&I Retrofit Program Evaluation.” (* Megdal et al. 1995a).

The use of the trend variable did make a very large change in the savings estimate in the direction predicted by our initial theoretical economic analysis. The demand model without the trend variable showed a realization rate of 105%, while the model corrected for autocorrelation showed a realization rate of 55%. Here again, the importance of testing and then correcting for error assumption issues is shown. This work also points to the need to consider the issues associated with estimating realization rates when strong economic trends are occurring.

Another study shows the need for, and the use of, numerous regression diagnostics, corrections, critical examination, and methods discussed here and above.⁸³ The examinations conducted, corrections employed and comparisons made to assess reasonableness are too numerous to list here. Instead, a brief summary is provided. The evaluators found “the diagnostics for the time-series/cross-sectional models complex, but crucial.” The format of their modeling immediately allowed an elimination of the first-order correlation in the pooled model. They used a comparison group created from several assessments to ensure a proper one that could control for exogenous factors. Engineering priors were created to provide an engineering estimate for evaporative cooler use for each customer in each time period (based upon cooling degree days in their area and floor space). This added more prior information to the model (a significant improvement over using a tracking system annual estimate). They created a correction factor so that the gross savings estimate from the participant/non-participant billing analysis did not underestimate gross savings by incorporating free ridership (which was estimated separately using a three-option nested logit approach). They found one regression diagnostic (DFBETA, a measurement of the influence of a particular observation) to be particularly useful in their analysis. Another important correction was the use of time period fixed-effect dummies that otherwise indicated a substantial omitted variable bias.

The above example also points out an example of a pathway in the billing analysis prescriptive roadmap. The billing analysis roadmap shows that one of the methods has a side-arrow going from billing analysis methods to engineering analysis methods and back again to billing analysis. This method was used in the Samiullah example.⁸⁴ For this evaluation, new engineering priors were developed to be more accurate by being time and customer specific.

Another study by Coito and Barnes conducted a more extensive engineering adjustment to engineering priors and one that was specifically done in relationship to regression diagnostics.⁸⁵ This study used a three-stage process for the impact evaluation. The first stage was a preliminary regression analysis that used telephone survey data and program tracking data along with the billing data. Regression diagnostics performed on this work were used to identify outliers that would then be the focus of on-site surveys and engineering analysis. The on-site surveys were used to (1) refine telephone survey data

⁸³ “Bells, Whistles, and Common Sense: Billing Analysis of a Residential HVAC Rebate Program.” (* Samiullah et al. 1996).

⁸⁴ (* Samiullah et al. 1996).

⁸⁵ “Improving Billing Analysis Results Using On-site Follow-up Surveys.” (* Coito and Barnes 1996).

and (2) explicitly quantify non-program changes that altered electricity consumption and potentially masked program energy savings. The final regression model was a SAE Change Model with a vector of explanatory variables explaining change in consumption. The findings of this staged approach of engineering and billing analysis indicated that (1) the root mean square error was much lower than expected, (2) more observations were able to be included into a meaningful regression, (3) the coefficient for the on-site survey-based quantification of non-program factors provided evidence verifying the likely accuracy of this work, (4) the model was more stable with regard to the inclusion or exclusion of outliers, and (5) the program realization rate rose from 63% to 84%.

Billing analysis should examine outliers and their influence on the results, as shown in the above example. There are a variety of methods that can be used to treat outliers. These can include using a trimmed mean, weighting (bi-weight mean), and median, among others, as seen in Pigg and Blasnick.⁸⁶ Another potential method includes bootstrapping techniques.⁸⁷

Summary Guidelines

The primary purpose of impact evaluation is to obtain the most accurate and unbiased estimate of energy and demand savings due to the program. A high quality billing analysis would produce these estimates in a manner that is defensible on the basis of current academic standards.

There are many types of errors, invalid assumptions, and threats to validity that can appear and may be very important to the final estimate of savings. A professional practice of examining the analysis with a critical eye and *knowing* its data are important to a quality evaluation.

The best evaluations will include reasonable disclosure of potential biases, specification problems, and data limitations.

Engineering Analysis

Engineering methods use basic rules of physics to calculate estimates of energy and demand savings. The technical information required as inputs to engineering models generally come from manufacturers, research studies, and other general references combined with assumed or measured equipment operating characteristics.

In order to estimate savings via engineering methods, one must establish a **baseline** from which to compare the energy consumption and demand of facilities included in the

⁸⁶ “Dealing with Outliers in Impact Evaluations Based on Billing Data.” (Pigg and Blasnik 1993).

⁸⁷ “The Treatment of Outliers and Influential Observations in Regression-Based Impact Evaluation.” (Schutte and Violette 1994). Another work that used bootstrapping and outlier correction, in this case mean absolute deviation estimator (MAD), can be found in “Billing Data Analysis of the C&I Sector: Application of Monthly Panel Models.” (Ozog et al. 1995a).

evaluation.⁸⁸ The program baseline may require a definition of the pre-program equipment or building characteristics and operations, as well as an estimate or measurement of pre-program energy consumption. The program baseline may consist of the following:

- For early equipment replacement (retrofit) programs, the pre-existing and still-functioning equipment replaced during program participation defines the baseline. Pre-program energy consumption may be adjusted to reflect changes in equipment or building operations not related to the program.
- For equipment that is being replaced at the end of its useful life (i.e., in all situations where the customer would have been replacing the equipment in the absence of the program), standard-efficiency new equipment defines the baseline. The program's purpose in these cases is to induce customers to do the replacement with a higher-efficiency alternative than they would have selected in the absence of the program.
- For operations and maintenance (O&M) programs (such as air conditioning tune-up or retro-commissioning programs), the existing condition of the equipment or existing O&M procedures define the baseline. Pre-program energy consumption may also be adjusted to reflect changes in equipment or building operations not related to the program.
- For new construction programs, the California Energy Efficiency Standards (Title 24) that define minimum standards for new construction, and Appliance Efficiency Regulations (Title 20) are used as the baseline. For program attributes not addressed by Title 20 or Title 24 (such as grocery store refrigeration systems), a "common practice" study may be conducted to establish the program baseline. Pre-program energy consumption data cannot be measured in new construction since the building does not exist. The energy implications of the baseline building characteristics are generally calculated using a building energy simulation program.

The Engineering Analysis Roadmap is presented in Figure 6.3. The various engineering analysis methods, literature references, and examples are presented in this section. The net-to-gross methods, issues, and references are provided in a later section within this chapter.

⁸⁸ Engineering methods used to date estimate savings as the difference between baseline and post-program usage. This requires an estimation of baseline. Econometric methods can estimate baseline and post-program and use the difference to estimate energy savings. However, there are econometric methods that can directly estimate savings rather than this more indirect difference approach.

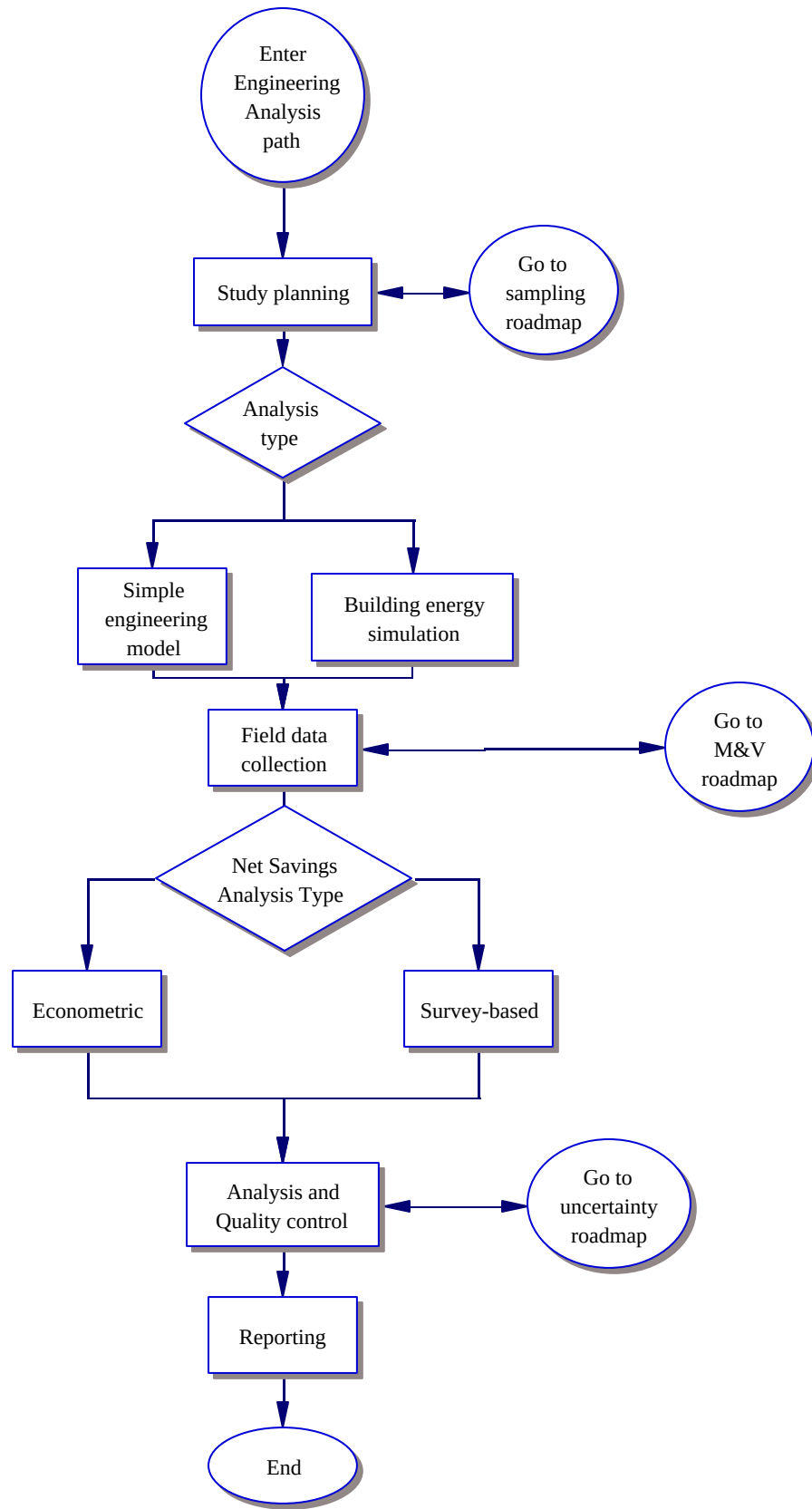


Figure 6.3: Engineering Analysis Roadmap

Engineering methods can be divided into two basic categories.

- Simple Engineering Model
- Building Energy Simulation Model

Simple Engineering Models

Simple engineering models and algorithms are typically straightforward equations for calculating energy and demand impacts of non-weather dependent energy efficiency measures, such as energy efficient lighting, appliances, motors, cooking equipment, etc. Simple engineering models are generally not used for weather dependent measures such as building envelope and HVAC measures; these measures are generally analyzed using building energy simulation models. HVAC interactions with internal loads such as lighting in conditioned spaces are generally calculated from a building energy simulation program. However, virtually all measures can be estimated using a simple engineering model, provided weather or load dependant parameters are developed by a more sophisticated method. Simple models also require knowledge of the efficiency and baseline operating conditions for the equipment that existed before the energy efficiency investment.

Simple engineering models are often incorporated into the program tracking system to provide an initial savings estimate based on program accomplishments and building or project characteristics data collected during the course of the project. These estimates may also be used as part of a statistical billing analysis. See Chapter 13, Sampling, for more information on the use of program tracking data in the context of sample design.

The general form of the gross annual energy savings equation is as follows:⁸⁹

$$\Delta kWh_{gross} = units \times \left[\left(\frac{kW}{unit} \times RLF \right)_{base} - \left(\frac{kW}{unit} \times RLF \right)_{ee} \right] \times FLH \times (1 + HVAC_c)$$

⁸⁹ Subscripts in the engineering equations are defined as:

base = baseline measure or technology
 ee = energy efficient or enhanced measure or technology
 s = summer
 w = winter
 d = demand
 c = energy consumption
 pk = peak
 heat = heating system
 cool = cooling system

Annual or seasonal average values will be denoted with a bar, such as $\bar{\eta}$, $\overline{EER}$, and so on.

where:

ΔkWh_{gross} = gross annual energy savings
 $units$ = number of technology units installed under the program
 $\frac{kW}{unit}$ = nameplate power consumption of baseline and energy efficient units
 RLF = rated load factor
 FLH = full load operating hours
 $HVAC_c$ = HVAC system interaction factor for annual energy consumption

The general form of the gross summer demand savings equation is as follows:

$$\Delta kW_{gross,s} = units \times \left[\left(\frac{kW}{unit} \times RLF \times DF_s \right)_{base} - \left(\frac{kW}{unit} \times RLF \times DF_s \right)_{ee} \right] \times CF_s \times (1 + HVAC_{d,s})$$

where:

$\Delta kW_{gross,s}$ = summer gross coincident demand savings
 $units$ = number of technology units installed under the program
 $\frac{kW}{unit}$ = nameplate power consumption of baseline and energy efficient units
 RLF = rated load factor
 DF_s = summer demand diversity factor
 CF_s = summer coincidence factor
 $HVAC_{d,s}$ = HVAC system interaction factor at utility summer peak hour

Terms for Simple Engineering Equations

Units. Units refer to the number of individual technology units installed under the program, normalized to some convenient quantity, such as lighting fixtures, motor horsepower, tons of air conditioning, square feet of insulation, etc.

KW/unit. This factor refers to the connected or "nameplate" load of the baseline or energy efficient technology, normalized to some convenient unit of measure. It is often referred to as the **unit demand**. The normalization parameter used should be consistent with the units defined above. In some cases, the number of units of technology installed at a particular location is different for baseline and energy efficient technologies (e.g., lamps or fixtures in overlit spaces may be removed during an energy efficiency upgrade). The value used for the baseline unit demand should include the power of all baseline technologies replaced by each unit of energy efficient technology.

The expression of the unit demand will depend on the type of technology evaluated. Lighting measures are characterized simply by the fixture input power. Motors are generally characterized by their nameplate horsepower, rather than input power. The unit of measure used in horsepower, and motor kW per horsepower is calculated as follows:

$$kW/hp = \frac{0.746}{\eta}$$

where:

hp = nameplate motor horsepower
 η = motor efficiency (unitless)
 0.746 = conversion factor (kW/hp)

Energy efficient cooling equipment is characterized in a similar fashion. The measure quantity data is expressed in terms of tons of cooling capacity. The unit demand is expressed in terms of kW per ton of cooling, and is calculated as follows:

$$kW/ton = \frac{12}{EER}$$

where:

EER = energy efficiency ratio of the cooling equipment (Btu/Wh)⁹⁰
 12 = conversion factor (kBtu/ton-hr)

For measures involving increased levels of insulation, such as wall insulation or water heater insulation, the unit demand is calculated from the thermal load and the efficiency of the equipment meeting the load. When the units are expressed in terms of insulation surface area, the unit demand is:

$$kW/SF = \frac{U \times \Delta T}{\eta \times 3413}$$

where:

U = overall heat transfer coefficient (Btu/hr-ft²-°F)
 ΔT = temperature difference (°F)
 η = efficiency of energy conversion device (unitless)
 3413 = conversion factor (Btu/kWh)

For processes involving fluid flow, such as water heating and air infiltration, electricity demand is also calculated from the thermal load and the efficiency of the equipment meeting the load:

⁹⁰ The energy efficiency ratio (EER) as published by air conditioning equipment manufacturers refers to the efficiency of the unit under specific outdoor and equipment entering air conditions, which may not correspond to the peak or average load conditions in a particular climate region. Under peak conditions in the inland regions of California, outdoor temperatures are generally hotter and entering air conditions are generally drier than the standard conditions used to evaluate the equipment EER. Residential equipment is often rated with a seasonal energy efficiency ratio (SEER), which is intended to be representative of the average efficiency of an air conditioner over a national average cooling season. EER data may not be available for residential units, although the Consortium for Energy Efficiency (CEE) maintains a database of EER data for residential units. See www.cee1.org.

$$kW = \frac{\dot{m} \times c_p \times \Delta T}{\eta \times 3413}$$

where:

$\dot{m}$	=	fluid mass flow rate (lb/hr)
c_p	=	fluid specific heat (Btu/lb-°F)
ΔT	=	temperature difference (°F)
η	=	efficiency of energy conversion device (unitless)
3413	=	conversion factor (Btu/kWh)

Rated load factor. The rated load factor is defined as the ratio of the maximum operating power of a measure to its nameplate power. It is used to correct for discrepancies in the nameplate power input or capacity of a piece of equipment relative to its actual peak operating power or capacity. These discrepancies can come from:

- equipment over-sizing due to conservative design approaches used by architects and engineers
- margins of safety applied by manufacturers to the nameplate power rating

In the context of appliances and other plug loads, the rated load factor is also referred to as the usage factor.⁹¹

Demand diversity factor. The demand diversity factor is used to account for the fact that not all measures in all buildings are drawing full power at the same time. The demand diversity factor is defined as the peak demand of a population of units to the peak demand of an individual unit. It is estimated by comparing the peak demand obtained from load research or end use metering studies to the connected load inventory and rated load factor for all applicable loads.

$$DF = \frac{kW_{pop}}{\sum_{i=1}^n (kW_i \times RLF_i)}$$

where:

kW_{pop}	=	peak demand of population of units
kW_i	=	nameplate rating of unit i
RLF_i	=	rated load factor for unit I

⁹¹ *The ASHRAE Handbook of Fundamentals*, Chapter 29 Nonresidential Cooling and Heating Load Calculation Procedures refers to the ratio of the nameplate load to the actual running load as the usage factor. (ASHRAE 2001b).

Coincidence factor. The coincidence factor is used to account for the fact that peak measure savings may not be coincident with utility peak demands. The coincidence factor is defined as the portion of the technology, system, or end use demand reduction that is coincident with the system peak.

Full load hours. Full load hours are a measure of the annual equipment operating hours, and are used to calculate annual energy savings. Equipment full load hours are defined as the total annual energy consumption divided by the equipment peak demand.

$$FLH_{eq} = \frac{kWh}{kW_{pk}}$$

For equipment servicing loads that do not vary with time (e.g., a motor driving a constant load), full load hours are simply equal to the operating hours. However, the power requirements of many types of equipment (e.g., air conditioning) vary during their operating hours; thus, the equipment full load hours are equal to the ratio of the hourly load to the maximum hourly load summed over all hours of operation.

Equipment full load hour data are usually obtained from end use metering or load research studies. These data are a function of the system or process load requirements, **and** the efficiency of the equipment meeting those loads. When analyzing DSM measures with different efficiency characteristics than their respective baseline technologies, it is often useful to define the **process** full load hours. When such equipment is replaced, the process load is not affected, yet the equipment full load hours may change. The process full load hours are related to the equipment full load hours by:

$$FLH_{proc} = FLH_{eq} \times \frac{\bar{\eta}}{\eta_{pk}}$$

where:

FLH_{eq} = equipment full load hours (calculated from load research or end use metering studies)

$\bar{\eta}$ = average equipment efficiency

η_{pk} = equipment efficiency under peak process load conditions

For example, the annual cooling load imposed on an air conditioner can be defined as the annual cooling load hours (CLH):

$$CLH = \frac{\text{Annual Cooling Load (Btu)}}{\text{Peak Cooling Load (Btu/hr)}}$$

Air conditioning equipment equivalent full load hours are defined as:

$$EFLH = \frac{\text{Annual Cooling Energy (kWh)}}{\text{Peak Cooling Demand (kW)}}$$

Cooling load hours and equivalent full load hours are related by:

$$CLH = EFLH \times \frac{\overline{EER}}{EER_{pk}}$$

where:

$$\begin{aligned} \overline{EER} &= \text{seasonal average cooling efficiency} \\ EER_{pk} &= \text{cooling efficiency under peak cooling load conditions} \end{aligned}$$

For many high-efficiency cooling machines, the ratio of the average cooling efficiency to the peak cooling efficiency is different than the baseline units replaced by the program. The process full load hours (CLH) remain constant (assuming the building has not been altered in any other way), but the equipment full load hours (EFLH) change as the ratio of the average to peak efficiency changes.

HVAC interaction factors. Certain measures, such as energy efficient lighting measures, directly influence lighting energy consumption and indirectly influence HVAC energy consumption. In certain climates, energy efficient lighting systems can reduce building internal heat gain, thus reducing cooling loads and increasing heating loads. Similar effects are produced by energy efficient appliances, cooking equipment and motors located in conditioned spaces. HVAC interaction factors can be defined for measures that have secondary influences on HVAC system energy. Separate HVAC interaction factors are defined for demand and energy savings calculations as follows:

- The HVAC interaction factor for **demand** is defined as the change in the HVAC demand per unit change in measure demand during the utility peak hour. Two factors are defined; the summer HVAC interaction factor generally considers the change in cooling demand, and winter HVAC interaction factor generally considers the change in heating demand.
- Similarly, the HVAC interaction factor for annual energy **consumption** is defined as the change in the annual HVAC electricity consumption per unit change in the annual measure electricity consumption. A single factor is defined which considers measure interactions with the annual heating, cooling, and/or fan electricity consumption.

There are many complicated interactions between internal gains, shell heat gains, thermal mass effects, HVAC system efficiency, and HVAC system control. Thus, HVAC interactions are best investigated with building energy simulation programs and for particular climate regions. Hourly simulation programs are best suited to calculating demand interaction factors. HVAC interactions are influenced by a number of factors, such as:

- Climate

- Relationship of building internal and solar loads to the overall envelope heat losses (balance point temperature)
- Coupling of internal load to HVAC load
- Building thermal mass
- HVAC system type
- HVAC system fuel
- HVAC system efficiency
- HVAC system controls

Thus, it is not appropriate to use a single HVAC interaction factor for all buildings, measures, and climates. HVAC interactions should be investigated over the range of climates and building characteristics listed above.

Simple engineering models can be calibrated through a combination of in-field metering and enhanced engineering analysis. Data collection is often carried out in a statistical framework, where the data collection and/or metering are conducted at a sample of sites. The results of the data collection are compared to the uncalibrated engineering estimate, and an adjustment factor is calculated.⁹² The adjustment factor is then applied to the engineering model results for all participants. If the data collection is carried out according to a reasonable statistical sampling plan, tests of statistical validity can be applied to the results.

Building Energy Simulation Models

Building energy simulation models are computer programs that use mathematical representations of important energy and control processes in an attempt to realistically simulate the thermal and energy systems in a building. Energy calculations are carried out on an hourly or sub-hourly basis for a selected time period or more commonly for an entire year based on typical weather data for the selected building site. The programs are made up of a collection of mathematical models of building components, such as windows, wall sections, and HVAC equipment. The individual component models are linked together to form a complete building simulation. The results predict the performance of the building structure and energy systems under given weather conditions at a selected geographic location. All building energy simulation programs have limitations that must be well-understood before applying the program to a particular energy estimation problem.⁹³ For example, most programs are limited to the simulation of common HVAC system types with a predetermined system configuration.

⁹² See *Quality Assurance Guidelines for Statistical and Engineering and Self-Report Methods for Estimating DSM Program Impacts* (* CADMAC 1998) for more information on calibrated engineering models.

⁹³ For more information on building energy simulation programs, see *State-of-the-Art Review: Whole Building, Building Envelope and HVAC Component and System Simulation and Design Tools*. (Jacobs and Henderson 2002).

Considerable latitude is given to the user with respect to describing system performance parameters, but the basic arrangement of the system component is fixed and defined by common practice in the building design and HVAC industries. This does not present a problem for most buildings and systems, but for complex custom HVAC configurations, the judgment and experience of the user is critical. In other situations, accurate weather data may not be available for the specific building site, and analogous weather data from a nearby location may produce some inaccuracy in the simulation results.

Below, a brief description of several building energy simulation programs commonly used in code compliance, evaluation, and building science research is presented.

DOE-2

DOE-2 is the most commonly used building energy simulation program for commercial buildings in the evaluation field. The DOE-2 is a **public domain** program, developed by the Lawrence Berkeley National Laboratory, with substantial funding from the U.S. Department of Energy. The first version of DOE-2 was released in 1979. The program has been modified and developed through 1999, when development was stopped in favor of the DOE-2.2 and EnergyPlus programs.

The final version of the DOE-2.1 program was DOE-2.1e. Within the DOE-2.1e version, there are 133 modifications or “releases.” A new release of the program may be stimulated by a series of “bug fixes,” or may include major new features. The California Energy Commission has adopted DOE-2.1e release 112 as the standard modeling program for non-residential Title 24 energy code compliance under the performance compliance option. This version is used in the EnergyPro Title 24 energy code compliance software.⁹⁴

The DOE-2.2 program is built on the DOE-2.1e program. DOE-2.2 includes a number of upgrades and changes, including improved simulation of window heat gains, duct losses,⁹⁵ heat recovery, and central plants in large buildings. DOE-2.3⁹⁶ includes the ability to model “refrigeration loops” that are made up of compressors, condensers, evaporators, and other components. This new functionality allows grocery store refrigeration and refrigerated warehouses to be simulated in detail.

EnergyPro

EnergyPro applies a user-friendly front end to the DOE-2.1e software engine to enable operation within a Windows environment. EnergyPro is certified by the California Energy Commission as a compliance tool for residential and non-residential buildings under California’s Title 24 Building Energy Efficiency Standards. As required by the CEC, EnergyPro calculates energy use for heating, cooling, and water heating for residential buildings; non-residential compliance calculations also account for lighting energy use. All forms necessary for California building permit submittal are output by EnergyPro. HVAC load calculations are available for both residential and non-residential

⁹⁴ (Gabel-Dodd/Energy Soft LLC 2001).

⁹⁵ The duct loss model in DOE-2.2 includes infiltration into the return side.

⁹⁶ Funding for the development of DOE-2.3 was provided by PG&E and SCE.

buildings. EnergyPro is used as a modeling tool for the California statewide non-residential new construction program (Savings by Design) and is approved for California ENERGY STAR[®] Home qualification.

Micropas

Micropas is a detailed energy simulation program that performs hourly calculations to estimate annual energy usage for heating, cooling, and water heating in residential buildings. Micropas is certified by the California Energy Commission as a compliance tool for California's residential Title 24 Energy Efficiency Standards. The program includes a load calculation for use in sizing heating and cooling equipment. All forms necessary for California building permit submittal are generated by Micropas. Micropas is approved for ENERGY STAR[®] Home qualification in California and nationwide.

Micropas has been in wide use in California since the early 1980s as a building energy code compliance tool and is growing in use elsewhere under the Model Energy Code. Up to fifteen thermal zones can be defined. Heating, cooling, natural ventilation and water heating systems for residential buildings are simulated. No detailed modeling of heating and cooling systems is provided - seasonal performance values are used.

EnergyPlus

EnergyPlus is the next-generation building energy simulation program. Released in April 2001, the program was developed jointly by Lawrence Berkeley National Laboratory, the University of Illinois, the U.S. Army Construction Engineering Research Laboratory, Oklahoma State University and others with support from the U.S. Department of Energy, Office of Building Technology, State and Community Programs.

EnergyPlus includes advanced simulation capabilities including time steps of less than an hour and modular HVAC systems simulation modules that are integrated with the building (zone) simulation. The architecture of EnergyPlus avoids several limitations inherent in the DOE-2 program. The response of the HVAC system to hourly zone conditions is greatly improved. HVAC systems of arbitrary design can be simulated, rather than relying on predefined templates. The program is seeing some use in the building science community, but has not seen wide use in the building design or evaluation community. Primary limitations are expanded data entry requirements, long run-times, and the lack of a user-friendly interface. Energy Plus is not certified for use as a Title 24 compliance tool in California.

The following table shows the relative applicability of each program to the analysis of energy efficiency measures in residential and commercial buildings. For more information, consult Jacobs and Henderson.⁹⁷

⁹⁷ *State-of-the-Art Review: Whole Building, Building Envelope and HVAC Component and System Simulation and Design Tools.* (Jacobs and Henderson 2002).

Table 6.4: Comparison of Simulation Program Capabilities

(● indicates robust capability, ○ indicates partial capability, blank indicates issue not addressed)

	MicroPas	DOE-2	Energy Plus
Air Leakage Sealing	●	●	●
CAD Interface		○ ¹	●
Code Compliance	●	○ ¹	
Cool Roofs	●	●	●
Cooling Towers/Fluid Coolers		●	●
Daylighting		●	●
Desiccant Dehumidification		●	●
Displacement Ventilation		○	●
Duct Leakage	●	● ²	
Economizers/Advanced Control		●	●
Exterior Shading	●	●	●
Ground Coupling	○	○	●
Heat Recovery Systems		●	●
High Performance Glass	○	●	●
IAQ			●
Improved Unitary Equipment	○	●	●
Interoperability			●
Lighting Design		○	○
Moisture Adsorption			●
Natural Ventilation	●	○	●
Passive Solar Heat	●	●	●
Photovoltaics			●
Radiant Barriers	●	○	●
Radiant Cool/Heat			●
Refrigerant Charge and Airflow	●	○	○
Refrigeration		●	
Simplified Inputs		○ ¹	
Sunspaces, Atria	●	●	●
System Sizing	●	●	●
Thermal Comfort		○	●
Thermal Mass	●	○	●
Thermostatic Expansion Valves	●	○	○
Variable Speed Pumping		● ³	●
VVT Systems		●	○
Zonal HVAC Systems	●	●	●

Note: 1. Simplified inputs, CAD interface and code compliance provided thorough third-party front-end programs.
 2. Full duct leakage modeling provided by DOE-2.2 only.
 3. Full variable speed pumping loop simulation provided by DOE-2.2 only.

Building energy simulation programs should be calibrated with billing data and/or monitored data to improve the accuracy of the model. Data collection for model calibration should be used to improve the inputs to the model and to verify the response of the model. Data collection to improve model inputs generally focuses on the model

parameters with the largest influence⁹⁸ on the results and the greatest uncertainty. See Chapter 7 on Measurement and Verification for more information on simulation model calibration.

Building Energy Simulation Model Validation

The U.S. Department of Energy, through the National Renewable Energy Laboratory, has been working with the International Energy Agency and the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) to develop standard methods of test for building energy analysis computer software. The Building Energy Simulation Tests (BESTEST) were developed to validate the response of building energy simulation programs. Several BESTEST protocols were designed to validate the simulation of building heating and cooling loads and the simulation of HVAC systems. ASHRAE recently published ANSI/ASHRAE Standard 140-2001, based on the BESTEST protocols.⁹⁹ DOE-2 and EnergyPlus have been validated under the ASHRAE 140-2001/BESTEST program.¹⁰⁰ Micropas has been validated under the building loads portion of the BESTEST procedure only.¹⁰¹

ASHRAE is considering an addendum to ASHRAE Standard 90.1 (*Energy Standard for Buildings Except Low-Rise Residential Buildings*) that would require any simulation program used to demonstrate ASHRAE 90.1 compliance to be validated using the tests contained in ASHRAE Standard 140. The authors believe that building energy simulation programs used for impact evaluation under this Framework should also be validated according to ASHRAE Standard 140. Simulation program validation should help identify weaknesses with the algorithms used within a program, thus removing an important source of bias from simulation results. Impact results developed from validated simulation programs that are calibrated to field data as described in Chapter 7 will provide the most reliable impact estimates.

Net-to-Gross Requirements and Method Options

The primary purposes of conducting a summative impact evaluation are (1) to provide reliable estimates of the energy and demand savings created by the program(s) for use in cost-effectiveness analysis, (2) to know how much resource can be depended upon, and (3) to incorporate savings in overall Public Goods Charge (PGC) fund estimates. In all of these cases what is really desired is the energy and demand savings induced by the program. In other words, the savings need to be “net” of what would have occurred in the absence of the program. Hence, we need to answer the question of what would participants (and non-participants) have done in the absence of the program that could

⁹⁸ Key model inputs are identified by calculating their influence coefficient. See “A Primer on the Use of Influence Coefficients in Building Simulation.” for more information on influence coefficients. (Spitler et al. 1989)

⁹⁹ (ASHRAE 2001).

¹⁰⁰ *EnergyPlus Testing with ANSI/ASHRAE Standard 140-2001 (BESTEST)*. (Henninger and Witte 2003).

¹⁰¹ BESTEST validation reported on the US DOE Building Energy Software Tools directory: http://www.eere.energy.gov/buildings/tools_directory/software/micropas.htm.

affect their energy use level. This sounds like a simple question. But 15-20 years experience has proven it to be a very difficult question to answer with any assurance as to the (unbiased) accuracy or precision of the answer.

These doubts in the quality of these estimates are not for lack of trying. The increasing study and sophistication of evaluation efforts to improve upon this capability throughout the 1990s was far more extensive in California than anywhere else. Thorough analysis by expert econometricians has repeatedly found problems in each new generation of “solutions.” Nevertheless, this section will summarize and reference several recent methods as possible choices.

Principles for Undertaking and Using Net-to-Gross Analysis

Often gross savings may be estimated by one method and net savings (the net-to-gross adjustment factor) estimated by another method. Sometimes, the analyses can be done hand-in-hand during billing analysis. Thus, a net-to-gross analysis should be conducted at the same time as gross savings impact evaluation. However, some of the methods can be costly, and evaluation resources can be limited. Accordingly, it is possible that net-to-gross analysis could be done less frequently than the gross savings impact evaluation, as long as a prior net-to-gross analysis is available at the time of the gross savings evaluation, and the program, market, and participants are suitably representative of those upon which the prior estimate is based. These factors need to be considered as evaluation resources are allocated in the evaluation planning process and as evaluations are designed to best meet the requirements of the evaluation goals and this Framework. (See Chapter 5, the Umbrella Roadmap, and the discussion of prioritizing evaluation resources. The allocation of resources between gross savings impact evaluation and net-to-gross analysis, both of which are used in the final net savings calculation, and the propagation of uncertainty calculations are discussed in Chapter 12, Uncertainty.)

The best available estimate of the net-to-gross ratio (NTGR) should always be used in program planning. The NTGR can be expected to vary depending upon the maturity of the equipment or service, the type of delivery in the program, the maturity of the program, and the customer sector. This means that the best NTGR estimate for program planning is the latest estimate for that program or a similar program. A deemed estimate based on one or more of these dimensions can be the fallback position for program planning.

The Framework recommends that deemed NTGR, however, never be used to report net savings within an evaluation (unless a net-to-gross method was attempted and no defensible estimate was able to be derived, i.e., the method failed and could not be corrected within the evaluation budget and priorities). The NTGR is an important input into cost-effectiveness analysis.¹⁰² As such, we need to ensure that evaluations produce the most reliable and precise estimates possible. An assessment of the different factors that help determine the NTGR (as listed above) shows that, though very unlikely, an unscrupulous implementer could “game” an environment that used deemed NTGRs. A

¹⁰² See the *Energy Efficiency Policy Manual, Version 2*. (* CPUC 2003), page 18.

program could be run that was lower cost by providing services to the most predisposed participants who would have taken, or are in the process of taking, the actions anyway (“free riders”) and yet claim upwardly biased net savings due to a high deemed NTGR. Though this is unlikely, it is prudent that this evaluation framework be designed to protect the use of PGC funds from this type of abuse.

Selecting an Analysis Method

The primary question is what would have participants done in the absence of the program. An obvious simple solution would be to compare the equipment installation rate of participants with the equipment installation rate among non-participants. However, if a program offers technical or financial services to anyone who wishes to participate, then those planning to take these actions would receive benefits with very little additional cost to them (just the program transaction/hassle costs). Hence, most people likely to adopt would participate in the program. (These people are called “free riders.”) Similarly, if someone was going to install energy efficiency measures later, they might be inclined to install these measures early to take advantage of the incentives. (These people are called “partial free riders.”) This means that the non-participant group is not a good comparison group for the participant group. There is a self-selection bias in who is a participant versus a non-participant, i.e., participation is not random. (This selectivity issue (among others) will continue to show up as a problem through the econometric methods.)

A simple comparison can demonstrate the nature of the selectivity bias. In this example, let’s assume that 12 percent of non-participants are found to install the same efficiency measure. A simple comparison might assume then that 12 percent of participants are free riders. However, recognize that some participants participated because they were going to make the installation anyway and wanted to receive “free incentives” for their planned activity, resulting in fewer non-participants in the non-participant pool installing measures. The proportion of “installing non-participants” is lower and the proportion of free riders is higher than would be indicated by comparing to the non-participant sample.

The impact evaluation roadmap categorizes NTGR methods very simply into those that are econometric methods (comparing participant and non-participants and adjusting for selectivity biases through econometric models) and those that are survey-based (asking participants what they would have done).¹⁰³ There are a variety of econometric methods that have been used with varying advantages and disadvantages. There are also a variety of more and less sophisticated survey-based methods that can include file review, review of letters in business decision process, payback exams and comparisons with other decision-making in the participating firms, engineering modeling, etc.

Econometric methods are sometimes considered the most accurate type of method, and they are preferred in California in situations where there are enough participants and comparable non-participants, and when the program is large enough to justify the

¹⁰³ At the same time, many of the econometric methods depend upon survey data from participants and non-participants to be used in the econometric model.

expense of the method. However, there are a few program types that make certain NTGR methods unfeasible. Programs with either a very small number of participants or non-participants (e.g., in sectors where there are only participants and former participants) or where comparability is a severe problem (e.g., programs that only work with large manufacturers) are not amenable to these econometric methods. This means that this small subset of programs needs to rely on a survey-based method. These could include participant surveys, vendor interviews, and/or record review methods.

An overview of the NTGR method options and examples of their use are provided in the discussions that follow.

Survey-Based Methods

Survey-based stated intentions, or “self-reports,” is a method of estimating free ridership by asking participants directly a series of questions on what they would have done in the absence of the program. Generally, the best use of this method has involved asking a series of questions with each question allowing a scale of responses. The initial question asks the participant, if the program had not existed, would they have installed the same equipment? The scale of responses for this question is whether they “definitely would have,” “probably would have,” “probably would not have,” or “definitely would not have” installed the energy efficient equipment. This scale, rather than a yes/no response, is thought by many to allow greater apparent accuracy in the estimate.

The program can also have an effect on the level of efficiency of the efficient equipment installed (when they say they would have installed the same efficient equipment without the program), when a participant installs efficient equipment, and how many they install. This is the element of partial free ridership. The continuum of free riders is presented in Figure 6.4.

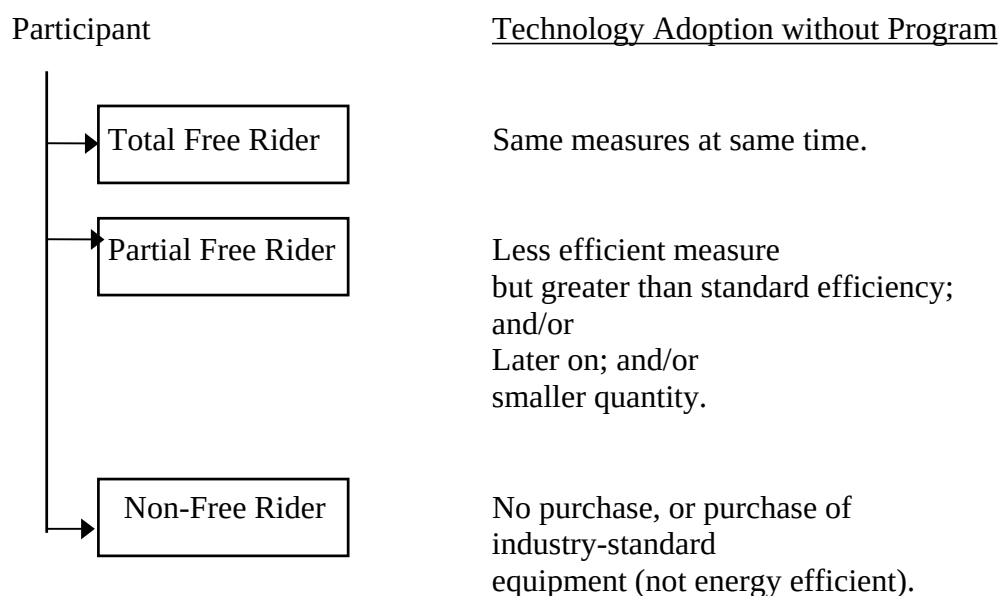


Figure 6.4: Continuum of Free Riders

The partial free ridership issue presented in Figure 6.4 is generally addressed through additional questions in a series of questions about stated intentions that are contingent on the response in the primary question, i.e., the questions are only asked of those that probably or definitely would have taken the actions. Examples of partial free ridership questions include the following:

1. If you had installed measures without the program, do you think the equipment would have been: “definitely as efficient,” “probably as efficient,” “probably not as efficient,” or “definitely not as efficient.”
2. If you had installed without the program, do you think the [measure] would have been installed at the same time?
3. If not installed at the same time, when would it have been installed? (Such as: earlier or later, or within three months, six months, next year, etc.)
4. Without the program, would you say the quantity of high efficiency measures you purchased would have been: “definitely greater,” “probably greater,” “the same,” “probably less,” “or definitely less.”

Each of the responses in the series of questions is assigned a probability for the expected net savings. These estimates are then combined (additively or multiplicatively) into a participant estimate. The participant estimates are subsequently averaged (or weighted averaged given expected savings) to calculate the overall free ridership estimate. The assignments of the probabilities are critical in the final outcome. At the same time, there is little evidence of what these should be and they are often assigned and justified given a logical argument. With this, however, a multiple number of different probability assignments have been shown to be justified and accepted by various evaluations and

regulators. However, this can make the comparability and reliability of survey-based estimates problematic.

The survey method is the most straight-forward method of free ridership estimation, and the lowest cost method. It does, however, have its disadvantages in potential bias and with accuracy. One method for improving this estimate is to add a consistency check question and adjust the individual's estimate accordingly. For example, asking the participant how important the incentive was in their decision to select the energy-efficiency equipment. If the answer from this question and the overall free ridership estimate from the prior questions are completely inconsistent, then the estimate for that participant is either modified, or additional questions are asked to clarify and obtain a better estimate for that participant, or the participant is removed from the analysis.¹⁰⁴

Early methods to handle responses of "Don't Know," missing data, and inconsistent answers involved assuming a 35% or 50% free ridership rate for these participants (as they might be less likely to have taken actions if they hadn't thought about it or made opposing reactions). These methods, however, were found to create a centrality tendency in the overall free ridership estimate, i.e., driving it towards 35% or 50%.

For example, a study in 1993 that created a standardized survey method for the New York utilities used these replacement assumptions. Then in 1997 the centrality bias was found to create NTGRs that were not reasonable in a study conducted for the Long Island Lighting Company (LILCo). This work created a NTGR estimate based upon removing the large number of inconsistent respondents from the analysis, rather than the prior procedure of substituting a central estimate for them, a revision to the standard method approved by the New York Public Utility Commission.

Due to this centrality tendency, the Framework recommends dropping respondents with inconsistent responses rather than assuming central estimates for them. Nevertheless, using additional survey questions to clarify and provide better estimates for these participants would be significantly preferable to losing these participants in the NTGR analysis. (Computerized surveys could be designed to check for the inconsistency and then add appropriate survey questions as needed to try and salvage a reasonable estimate for that respondent.)

Other questions have been used to help provide backup or adjustments for the free ridership estimates. For example, one can evaluate the decision-making process and identify where the participant was in that process when they learned about the program. This method is particular useful for customers that have difficult and time-consuming processes, such as government and large commercial customers. There have also been other studies examining information dissemination, other decision priorities, and general criteria for commercial investments.¹⁰⁵

¹⁰⁴ *DSM Free Ridership Study* (Cambridge Systematics Inc. and Freeman Sullivan and Company 1994).

¹⁰⁵ An example of using alternative questions and comparing their responses can be found in: *DSM Free Ridership Study* (Cambridge Systematics Inc. and Freeman Sullivan and Company 1994); and "FreeRidership Estimation in the New Construction DSM Market." (* Tolkin and Reed 1993).

Generally, the simplest and lowest cost NTGR method is using the survey-based stated intentions method with a telephone survey for data gathering. Unfortunately, there has been significant research that this method can easily provide biased results.¹⁰⁶ One type of bias is that respondents can overestimate what they would have done, either because they see that response as socially desirable or because respondents do “intend to” save energy, but there are many things people can intend to do that never happen or are postponed significantly. These problems can also be increased depending upon the wording of the survey questionnaire, location of the survey questions (response effects’ bias), and the length of time involved in recall (the telescoping problem).¹⁰⁷ As a result, one study concluded that the survey-based method is most likely to be biased.

“This is the most important component of an estimate’s validity but can not be measured from statistical confidence intervals. Inexperienced analysts can easily be deceived that their answer is quite accurate when their survey instrument or assignment strategy is biased.”¹⁰⁸

Survey-based stated intentions is a classic case of potential bias in measurement, non-random measurement error or an issue with construct validity (i.e., we are not properly measuring the construct – free ridership - we are trying to measure). Coming at the question of what the participant would have done in the absence of the program from a variety of different perspectives (directly asking, decision-making criteria, where they were in the process, etc.) and assessing these together is one way the survey methods have used to triangulate on the correct construct.

A few studies have specifically examined how we can improve the validity of our surveys and minimize their potential biases. One study found that strict, factual, and clearly defined terms appear to minimize social desirability bias.¹⁰⁹ Another study tested for the ability of survey design to improve the accuracy of self-reports by using graphics in WebTV surveys and assessing false positives and false negatives. The evaluator reported that there were circumstances when the graphics were quite useful and others where they were difficult to employ, or without enough advantage to justify other issues that must be considered e.g., depending upon WebTV surveys in cases where that might not be appropriate.¹¹⁰

¹⁰⁶ *Net-to-Gross Ratios for PG&E’s CIA Rebate Program* (Xenergy and Cambridge Systematics Inc. 1993); and *DSM Free Ridership Study* (Cambridge Systematics Inc. and Freeman Sullivan and Company 1994).

¹⁰⁷ The terms provided in parentheses are from research and survey design. Experience with these elements and this literature are necessary for designing a quality evaluation study that uses surveys for data gathering.

¹⁰⁸ (Cambridge Systematics Inc. and Freeman Sullivan and Company 1994).

¹⁰⁹ “Sure You Do. Un-Huh’: Improving the Accuracy of Self-Reported Efficiency Actions.” (* McRae 2002).

¹¹⁰ “Effects of Using WebTV Graphics on the Accuracy of Self-Reports.” (* Peters 2002).

We can never know the “true” free ridership rate.¹¹¹ Survey methodology can provide a measure of relative precision that appears good based upon large sample sizes.¹¹² This is not the same thing as being confident that the answer is accurate. (See the Chapter on Uncertainty for a discussion of precision versus bias.)

Finally, survey-based methods may be appropriate for smaller programs when this is the only reasonable (affordable) NTGR method to use.

Given the ease with which the survey instrument and assignments can create varied NTGR estimates, and therefore, the potential for this to be “gamed” to obtain underestimated NTGRs, the Framework recommends that the CPUC consider supporting an overarching evaluation study to develop a standardized survey method for use in California. A carefully constructed survey with checks and triangulation methods might offer a more “accurate” measurement of a particular program’s free ridership, particularly given differences across programs. A standardized survey method may not provide greater “accuracy.” However, it can allow a free ridership survey method to be cost-effective for small programs to undertake, and it ensures consistency of measurement technique and application across these programs. This can help facilitate meaningful comparisons across programs for portfolio analysis, reporting, and planning purposes.

Massachusetts regulators recently asked for a study to derive a standardized free ridership survey method to be used by all the Massachusetts utilities for all their program evaluations. This study was completed in mid-2003, and the first program evaluations using it are just being completed.¹¹³ The CPUC may want to use this study or conduct a similar study for California programs.

There are two improvements to the Massachusetts study that the Framework recommends for consideration if a standardized method were to be developed for use in California. First, a method needs to be developed for using contingency questions and for dropping responses from inconsistent respondents from the analysis if the inconsistency cannot be corrected (rather than substituting a central measurement). Second, testing and analysis is needed to calibrate the survey against a mix of econometric methods as discussed below.

¹¹¹ There have been comparison studies between survey-based methods and nested logit methods that pointed to the differences as illustrating the bias in the survey-based methods. However, bias could also exist in the nested logit method and concerns with this method have been discussed in later work. These earlier comparisons can be found in *Net-to-Gross Ratios for PG&E’s CIA Rebate Program*, (Xenergy and Cambridge Systematics Inc. 1993) and *DSM Free Ridership Study* (Cambridge Systematics Inc. and Freeman Sullivan and Company 1994).

¹¹² The confidence level is providing precision that the sample is measuring the same answer as would be received if the full population had responded. The problem is not one of sampling but of construct validity. That is, do the specific questions asked and the response assignments, weights, and combining of question accurately measure what would have occurred in the absence of the program.

¹¹³ *Standardized Methods for Free Ridership and Spillover Evaluation – Task 5 Final Report (Revised)*. (* PA Consulting Group Inc. 2003).

Examples of Survey-Based Methods with Record Review and/or Engineering Components

Asking a series of questions on what participants would have done in the absence of the program is problematic in complex and dynamic sectors, such as the non-residential new construction market, industrial projects, and large customized projects. In some cases, such as with new construction, the decision process involves many steps, there is no comparison point before action is taken, and it is difficult for the participant to imagine what “would have occurred.” In other cases, participants are hesitant to provide unbiased information given perceptions about their firm, confidentiality on how investment decisions are made, or perceptions of how their answers might affect their relationship with the program, the utility, or future funding of the program. This complication has been dealt with through careful, and sometimes customized, interview design and analysis methodologies.

For example, Goldberg and Scheuermann present work where an interview script was designed to be administered by an evaluation engineer familiar with the project and able to apply the inquiry with a discussion of project details on-site for large and custom projects. The project detail discussions can help the participant focus in on each component of the decision for a complex project or decision-making process.¹¹⁴

In another study, information was gathered for a net-to-gross analysis of an industrial program which included a review of program files, operations staff surveys that were used to create customized free ridership questions for a decision-maker survey, a decision-maker survey, an on-site survey, and a vendor survey. This multi-step process allowed customized questions to be created to better target and disaggregate the complex projects and the decision-making being examined.¹¹⁵

There are advantages of using multiple methods as either back-up or triangulation to obtain the most defensible NTGR estimate, particularly in cases with complex projects and decision-making. Differing circumstances can also significantly affect the accuracy of a NTGR methodology.

For example, in an evaluation of California’s non-residential new construction program, Savings By Design, two different methods were used. One method was based on-site visit data, baseline study, and engineering models, creating an engineering-based difference of differences between participant and non-participants (i.e., comparing participants versus non-participants on a baseline model estimate versus an as-built model estimate). The second method included self-reported intentions soon after construction for both participant and non-participants and combined them with engineering modeling. The evaluators noted that increasing market effects and potential impacts from the California energy crisis might have been showing up as non-participant spillover. With this in mind, they reported a problem with the difference of differences methodology. This method left out non-participant spillover and underestimated the

¹¹⁴ “Gross and Net Savings Analysis for Unique Projects.” (Goldberg and Scheuermann 1997).

¹¹⁵ “A Combined Engineering and Decision-Analysis Methodology for Evaluating Spillover and FreeRidership in PG&E’s 1995 Industrial Energy Efficiency Program.” (Reid et al. 1997).

NTGR from free ridership. The second method, using self-reports and engineering data, provided a more defensible (where anyone might believe) NTGR in the presence of non-participant spillover for the second year.¹¹⁶

Econometric-Based Methods

The purpose of the econometric-based methods for estimating the NTGR is generally to use a comparison with non-participants and control for differences and the selectivity bias inherent in the participation decision. This section's discussion focuses upon selectivity bias. It is important to recognize that using a non-participant comparison group controls for many other occurrences other than whether the measure would have been adopted or not, such as: changes in the economy, behavioral trends, fads, and other historical changes.

Some impact billing analysis models attempted to measure net savings by performing a billing analysis on both participants and non-participants with one dummy variable to indicate the post-retrofit time period for both participants and non-participants (as a control for exogenous effects on both) and a second dummy variable for only participants during the post-retrofit period. The coefficient on participation was meant to measure "net savings." (This variable is zero for non-participants. This means that the participant post-retrofit change is allocated between the post-time period dummy where all the non-participant post data exist and the post-participation dummy.) This type of analysis does generally lower the savings estimated by the billing analysis compared to not including this selectivity correction (except in times of booming economic growth or due to historical issues). Yet, this does not mean that it is a reliable estimate of net savings. The post-period change for participants is pulled downward by the non-participant data being within that variable, but there is still a selectivity bias that exists. Given the selectivity bias in participation, the savings from this analysis is an overestimate of net savings (as the reduction in the post-period for non-participants is less than the participants would have seen as some greater proportion of participants would have adopted the actions).^{117, 118}

A common correction methodology for standard selectivity bias in econometrics is the Heckman correction factor by using an inverse Mills ratio. Heckman first proposed its use in 1976 in labor economics when examining the effect of education on wage equations with truncated data, i.e., a selectivity that if a laborer's wage to be received is lower than their "reservation wage" (the lowest wage for which they are willing to work), no wage would be observed for them.¹¹⁹ The method was then used in other areas

¹¹⁶ "Measuring Accomplishments of Energy Efficiency in California's Nonresidential New Construction Market." (* Brost et al. 2002).

¹¹⁷ An example of a simple participant/non-participant comparison adjustment factor that incorporates these issues can be seen in "Matching Methodologies and Program Types: An Evaluation Retrospective." (Conant and Schutte 1993).

¹¹⁸ This argument was presented in "Why Discrete-Continuous Billing Models Mis-Estimate Net Savings of DSM Programs." (* Paquette 1996).

¹¹⁹ "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models." (Heckmann 1976).

beginning with employment and job training. The method has also been refined and modified across a variety of applications.^{120, 121}

The Heckman method of using an inverse Mills ratio was then used in energy efficiency analysis. Yet, it has since been shown that the type of selectivity bias that these methods address are not the same as the participation selectivity bias found in energy efficiency program evaluations.¹²² For example, in the above original use of the correction factor, the selectivity is caused by truncated data; those without work have no wage data and are more likely to have the lowest levels of education. This logic is equivalent to the fact that customers more likely to install the measures are more likely to participate in the program (and they cannot participate in the program without installing). This is the primary selectivity issue discussed above. In addition, there is a second correlation of concern for the econometric modeling. Those with greater savings over what they would have done otherwise (those with lower free ridership percentages) are more likely to install and participate in the program. (It is this second factor that led to the use of the double inverse Mills ratio, as discussed below.)

A third method used to correct for selectivity is a two-stage regression where the first stage estimates the probability of participating in the program, which would then become a variable in the second stage that ran the billing analysis for net savings. Given that the probability to participate is a discrete choice (you either participate or not), a discrete choice analysis is needed to estimate this regression. The discrete choice model includes variables hypothesized to lead to the decision to participate in the program. The billing analysis then contains variables related to energy usage plus the probability of participation (estimated from the discrete choice analysis). There have been a number of impact evaluations estimating net savings through this method (termed discrete-continuous billing analysis).¹²³ This approach yields higher estimates of free rider savings and lower estimates of net savings than the basic participant/non-participant billing analysis, which is what would be expected with a correction for this type of selectivity bias. However, the discrete-continuous method only corrects for one of two decisions that may differ between participants and non-participants, the decision to participate. There may still be a selectivity bias due to differences in the decision to install the equipment. Hence, the method may be increasing non-random error and uncertainty.¹²⁴

Nested logit models have been used to estimate naturally occurring conservation investments or actions to derive net-to-gross ratios. The theoretical basis for this work pointed out that the nested logit could address the joint decision-making between

¹²⁰ *Time Series Techniques for Economists*. (Miller 1990).

¹²¹ "Sample Selection Bias as a Specification Error." (Heckmann 1979).

¹²² *Net Savings Estimation: An Analysis of Regression and Discrete Choice Approaches*. (* Xenergy 1996).

¹²³ An example can be found in "Combining Monitoring, Engineering Analysis, and Billing Analysis to Evaluate PG&E's Commercial Retrofit Incentive Program." (Buller et al. 1993).

¹²⁴ "Why Discrete-Continuous Billing Models Mis-Estimate Net Savings of DSM Programs." (* Paquette 1996).

installation and program participation. Nested logit attempted to solve this problem in that it jointly estimates the probability of installations and the probability of participation given that the measure will be adopted.¹²⁵ Nested multinomial logit (NML) analysis examines three possible outcomes: (a) do not install the measure, (b) install the measure and do not obtain a rebate, or (c) install the measure and obtain a rebate.¹²⁶ The nested logit model explicitly tests the joint decision-making between the decision to participate and the decision to adopt the technology.¹²⁷ This structure can be seen in Figure 6.5.

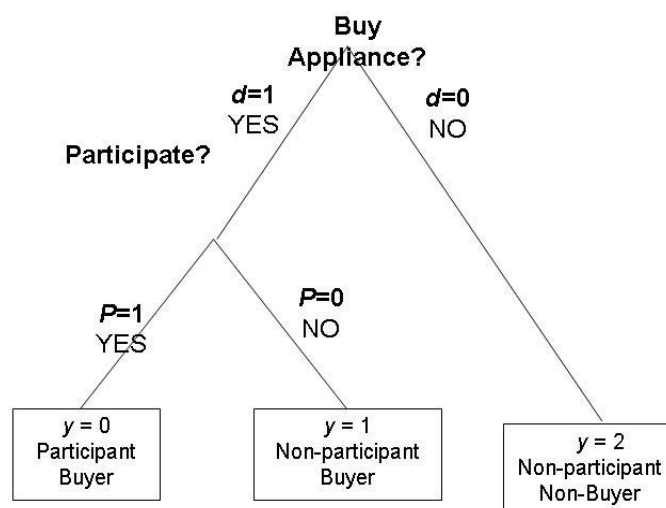


Figure 6.5: The Nested Logit Model for Adoption and Participation Decisions

Source: “Instrumented Decomposition: A Two-Stage Method for Estimating Net Savings.”¹²⁸
Similar graphics available in several of the other publications cited concerned use of the NML method.

An alternative econometric method developed was to use two inverse Mills ratios - the first was applied to the entire sample and a second was applied to only the participants. This method is known as the Double Mills Ratio method. It was developed to address the particular type of selectivity bias found with energy efficiency programs that is not

¹²⁵ “Estimation of Net Savings from Energy-Conservation Programs.” (Train 1994). Another comparison was made with Monte Carlo experiments with simulated data sponsored by the Wisconsin Center for Demand Side Research, “Is It Net or Not? A Simulation Study of Two Methods.” (Goldberg and Kademan 1995).

¹²⁶ “Estimation of Net Savings for Rebate Programs: A Three-Option Nested Logit Approach,” (* Train et al. 1994); “Modeling Technology Adoption and DSM Program Decisions.” (* Buller et al. 1994).

¹²⁷ See also: “Freerider and Freedriver Effects from a High-Efficiency Gas Furnace Program,” (* Seiden and Platis 1999); “Developing Confidence in Your Net-to-Gross Ratio Estimates,” (* Mast and Ignelzi 1996); “A Comparison of Two Net Analysis Methods Using Data from PG&E’s Nonresidential New Construction Program,” (* Heitfield et al. 1996); “Do Central Air Conditioner Rebates Encourage Adoption of Air Conditioning?” (* Samiullah et al. 2002).

¹²⁸ (Kandel 1999c).

properly addressed by the Heckman inverse Mills Ratio method.¹²⁹ A subsequent study tested this method in an evaluation of a commercial HVAC program,¹³⁰ and another study has used this method in a market effects study.¹³¹

In 1995/1996, a comparative analysis of the discrete choice method and the double inverse-Mills-ratio approach was conducted. The study confirmed the use of NML for rebate programs. It also found the double inverse Mills ratio could properly correct for self-selection bias given specific underlying population distributions. At the same time, the results and potential bias for using the double inverse Mills ratio could be worse than omitting the self-selection correction at all for some distributions.¹³²

The distribution issue for using a double inverse Mills ratio led to the development of “instrumented decomposition” (ID) as a method that did not need to assume any particular error distribution.¹³³ This method used NML to estimate the correlation of errors and then used an instrumental variable in a two-stage least-square model (2SLS). The ID method has been tested with data from an evaporator cooler program¹³⁴ and an air conditioning rebate program.¹³⁵ This method continues to be refined.

Most econometric methods for NTGR require survey information. The more they rely upon self-report type data, intentions, and psychographic data, the more they are likely to have some of the same measurement issues as the survey-based approach. The ID method seems promising. At the same time, there may be extensive data requirements for this method that might limit its applicability to programs that either have or can justify large data collection costs.

Summary of NTGR Method Options and Roadmap Choices

In summary, there are many econometric methods and survey methods that can be applied to the question of estimated NTGR. Most, if not all, contain issues with potential bias, limiting assumptions, or lack of outside verification and testing. On the other hand, many methods may provide defensible estimates for a wide variety of specific programs. The Framework recommends that a method be selected in the evaluation planning process that is consistent with the type of program, its program theory, knowledge of the

¹²⁹ A similar method was developed for a similar problem by James Heckman as published in “Alternative Methods for Evaluating the Impact of Interventions,” (Heckmann and Singer 1985). Thank you to Adrienne Kandel for provided us this information and other clarifications within this section.

¹³⁰ “How Many Mills Ratios Does it Take to Estimate Net Savings?” (* Randazzo et al. 1996).

¹³¹ “Self-Reports and Market Transformation: A Compelling New Approach.” (Cavalli et al. 1999).

¹³² *Net Savings Estimation: An Analysis of Regression and Discrete Choice Approaches.* (* Xenergy 1996)

¹³³ “Instrumented Decomposition: A Two-Stage Method for Estimating Net Savings.” (Kandel 1999c).

¹³⁴ “Evaporative Cooler Rebate Program Cuts Load Significantly, and May Overcome Class Barrier.” (Kandel 1999a).

¹³⁵ “Instrumented Decomposition: A New Method to Estimate the Net Energy Savings Caused by Efficient Appliance Rebate Programs.” (Kandel 1999b)

data collection effort that can be accomplished, and a critical approach to measurement and assessing potential problems.

Chapter 7: Measurement and Verification

Preface

The Umbrella Framework and Impact Evaluation (Chapters 5 and 6) provide some guidance on the role of measurement and verification (M&V) in the context of conducting an impact evaluation. This chapter provides more details on conducting M&V studies. As such, it can be useful for policy makers and implementers needing an overview of M&V methods and issues.

During the period of 1994 to 1997, California's measurement and evaluation protocols (sometimes referred to as the CADMAC Protocols) referenced the National Association of Energy Services Companies (NAESCO) standards for measurement and verification in Appendix H as a resource for M&V activities. The NAESCO protocols were the precursors to the International Performance Measurement and Verification Protocols (IPMVP)¹³⁶ established by the US Department of Energy in 1996. The California Energy Efficiency Policy Manual,¹³⁷ first published in 2001 and revised in 2003, references the IPMVP, directing evaluators to "reference the appropriate IPMVP option" and "state any deviations from [the] IPMVP approach" when developing evaluation plans for programs by the Public Goods Charge (PGC). This chapter provides additional information on the IPMVP, attempts to clarify the intent of the various IPMVP options as they relate to impact evaluation studies, and describes how these activities support the overall evaluation process.

M&V and Process Evaluation

Measurement and verification studies can provide important information to program implementers regarding measure installation verification, installation quality, manufacturing defects, measure use and operation, equipment maintenance procedures, and in-situ measure efficiency that can improve program design. While billing analysis is useful in providing energy savings, it generally does not provide measure-specific performance information or insights into why measure performance might be less than expected. Process evaluation and M&V activities conducted early in the program implementation cycle can provide important early feedback on potential program or measure problems to implementers at a point where effective corrective action can be taken. Thus, M&V can serve an important and complementary role with process evaluation studies.

¹³⁶ (International Performance Measurement and Verification Protocol 2001).

¹³⁷ (* CPUC 2003).

Skills Required for Measurement and Verification

Simple engineering equations and simple instrumentation such as run-time data loggers can be understood and used by most people with a general science background.¹³⁸ Large-scale data acquisition systems and complicated analysis techniques generally require someone with a graduate degree in mechanical or architectural engineering.

This chapter provides an overview of the methods used in M&V. It provides background and references for policy makers, evaluation managers, and evaluators new to this field but with some knowledge of M&V issues. It is not meant to provide a “how-to” manual. The discussion here is meant instead to provide an overview and a perspective on how M&V methods are applied in the task of evaluating energy efficiency programs. References to the literature are provided for more information on M&V.

Introduction and Key Issues

The purpose of M&V activities is to verify that measures promoted by a program were actually installed and to measure the gross electricity savings from the measure installation relative to some baseline pattern of use. M&V, in this context, refers to data collection, monitoring and analysis activities associated with the calculation of gross energy savings from **individual customer sites or projects**. Program level gross and net impacts will be guided by the impact evaluation chapter, which also contains a sampling section to guide decisions about individual site selection. The M&V chapter is a subset of the overall impact evaluation process. M&V activities will primarily support program impact evaluations. Protocols for data collection to support other evaluation activities will be covered in their respective sections.

Performance contracting programs, such as the Standard Performance Contract (SPC) program, require M&V as a component of the program offering. Incentives are paid to implementers based on measured energy savings. The M&V activities conducted for the program may also provide information for program evaluation.

The Role of the International Performance Measurement and Verification Protocol

The IPMVP is the basis of the M&V activities covered in this chapter. The IPMVP was first published in 1996 as the National Energy Management Verification Protocol (NEMVP) and has since gone through several revisions. The current version (as of this writing) was published in the last few years in three volumes:

- ***IPMVP Volume I: Concepts and Options for Determining Savings (2002 Edition)***. Volume I defines basic terminology useful in the M&V field. It defines general procedures to achieve a reliable and cost-effective determination of

¹³⁸ The Association of Energy Engineers (AEE) offers a certificate for a Certified Energy Manager (CEM). The material covered in the CEM program is good background for understanding energy engineering concepts addressed by measurement and verification.

savings. Verification of savings is conducted relative to the M&V Plan for the project. This volume is written for the general application of measuring and verifying the performance of projects improving energy or water efficiency in buildings and industrial plants. This volume is the primary reference for the Framework.

- ***IPMVP Volume II: Concepts and Practices for Improving Indoor Environmental Quality (2001 Edition)***. Volume II reviews indoor environmental quality issues as they may be influenced by an energy efficiency project. This volume focuses on measurement issues, project design, and implementation practices associated with maintaining acceptable indoor conditions for an energy efficiency project, while providing guidance on key related elements of M&V and energy performance contracts. Issues discussed in this volume have limited relevance to impact evaluation, but may be useful when quantifying indoor environmental quality improvements. See Chapter 11 on Non-Energy Effects for more information.
- ***IPMVP Volume III: Applications (due to be published in 2004)***. Volume III will provide guidance on application-specific M&V issues. Individual chapters on *Concepts and Options for Determining Energy Savings in New Construction* and *M&V for Renewable Energy Technologies* are available.

The IPMVP was originally designed as a protocol to verify energy-savings projects implemented by energy services companies (ESCOs) under a shared-savings type contract or a guaranteed savings contract. It has since found applications to a broad variety of energy and water conservation projects throughout the world, including recent work on the verification of energy-savings projects for climate change mitigation.¹³⁹ The CPUC's *Energy Efficiency Policy Manual, Version 2*¹⁴⁰ references the IPMVP protocols as a key component of an overall Evaluation, Measurement and Verification (EM&V) plan.

There are several related documents that reference and complement the IPMVP:

- ***Federal Energy Management Program (FEMP) M&V Guidelines***.¹⁴¹ This document is based on the 1997 version of the IPMVP, and provides more detailed guidance on the application of different M&V protocol options for specific energy conservation measures.

¹³⁹ IPMVP is listed as the preferred choice for monitoring and evaluating energy efficiency projects for climate change mitigation in *Guidelines for the Monitoring, Evaluation, Reporting, Verification and Certification of Energy-Efficiency Projects for Climate Change Mitigation*. (Vine and Sathaye 1999).

¹⁴⁰ (* CPUC 2003).

¹⁴¹ *Federal Energy Management Program (FEMP) M&V Guidelines: Measurement and Verification for Federal Energy Projects, Version 2.2*. (* FEMP 2000).

- **ASHRAE Guideline 14-2002 – Measurement of Energy and Demand Savings.**¹⁴² ASHRAE Guideline 14 provides additional detail on implementing M&V plans within the IPMVP framework, including detailed HVAC system examples. It also contains detailed information on quantifying and minimizing uncertainty in M&V plan design, data acquisitions system selection and design, sensor selection, sensor placement, and calibration.

Statistical Context for M&V Studies

M&V studies are generally not conducted at every individual customer site within a program. The selection of M&V study sites should be done according to a sampling plan that allows the individual site findings to be expanded to the full program with a quantifiable level of statistical precision. Similarly, it may be impractical and costly to apply M&V activities to all measures installed at a particular site. Measures within a particular site should be sampled in a manner that allows the results to be expanded to the full set of measures installed at the site with a quantifiable level of statistical precision. See Chapter 13, Sampling, for more information on sampling for M&V activities.

Upper Level Roadmap

The upper level of the M&V roadmap is shown in Figure 7.1. The appropriate M&V option is discussed in the next section.

¹⁴² Measurement of Energy and Demand Savings, Guideline 14. (* ASHRAE 2002).

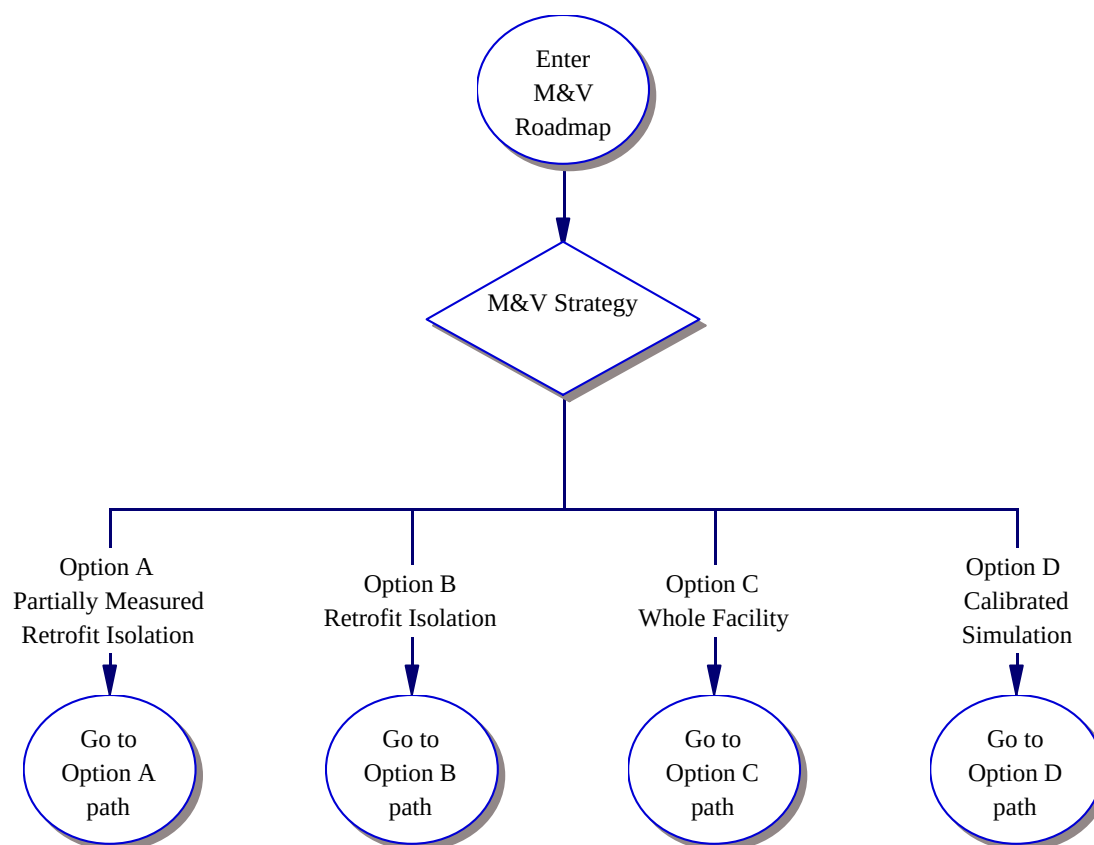


Figure 7.1: Upper Level M&V Roadmap

Measurement and Verification Options

This chapter provides several options for M&V studies. The options follow the terminology used in the IPMVP.

Option A - Partially Measured Retrofit Isolation. Savings under Option A are determined by partial field measurement of the energy use of the system(s) to which an energy conservation measure (ECM) was applied separate from the energy use of the rest of the facility. Measurements may be either short-term or continuous. Partial measurement means that some parameter(s) affecting the building's energy use may be stipulated, if the total impact of possible stipulation error(s) is not significant to the resultant savings. Careful review of ECM design and installation will ensure that stipulated values fairly represent the probable actual value. Stipulations should be shown in the M&V Plan, along with analysis of the significance of the error they may introduce. Savings are estimated from engineering calculations using short-term or continuous post-retrofit measurements and stipulations. A typical application of Option A is a lighting retrofit, where pre/post fixture watts are stipulated from a standard fixture wattage table, and operating hours are derived from short-term measurements of fixture run-time.

Option B - Retrofit Isolation. Savings are determined by field measurement of the energy use of the systems to which the ECM was applied separate from the energy use of the rest of the facility. Short-term or continuous measurements are taken throughout the

post-retrofit period. Savings are estimated from engineering calculations using short-term or continuous measurements. A typical application of Option B is a variable frequency drive applied to a constant speed pumping application. Pre-retrofit power consumption is measured with a hand-held power meter (or short-term metering to confirm constant power draw), and post-retrofit power consumption is short-term metered along with some relevant independent quantity (such as fluid or ambient temperature). The relationship between power and the independent variable is used to project long-term post-retrofit energy consumption from the short-term measurements.

Option C - Whole Facility. Savings are determined by measuring energy use at the whole-facility level. Short-term or continuous measurements are taken throughout the post-retrofit period and compared to 12 to 24 months of pre-retrofit data. Savings are estimated from an analysis of whole-facility utility meter or sub-meter data, using techniques ranging from simple comparison to regression analysis. This approach is very close in concept to a billing analysis, but may contain baseline adjustment factors that are specific to each building addressed under this option.

Option D - Calibrated Simulation. Savings are determined through simulation of the energy use of components or the whole facility. Simulation routines should be demonstrated to adequately model actual energy performance measured in the facility. This option usually requires considerable skill in calibrated simulation. Savings are estimated from energy use simulation, calibrated with hourly or monthly utility billing data, and/or end use metering.

Choosing the M&V Option

The M&V option employed will depend primarily on the impact analysis chosen and the overall precision goals for impact results as defined in the impact evaluation plan. In general, Option A is the least accurate and least costly option, whereas Option B or D, depending on the measure, are generally the most accurate and most costly options. All measures can be evaluated using Options A, B, and D, but the accuracy of the estimates provided under Option A decreases as the measure complexity increases. Option C is limited to projects where the expected savings exceeds the metered energy consumption by at least 10%.¹⁴³ A list of general considerations for selecting an M&V approach is shown in Table 7.1. Specific criteria for option selection are described in the section covering the details of each option.

¹⁴³ 10% is the minimum savings criterion established in ASHRAE Guideline 14. Depending on the variability of the data, a greater energy savings fraction may be required for a successful billing analysis.

Table 7.1: M&V Option Selection Criteria

Consideration	Option A	Option B	Option C	Option D
Ability to determine individual measure savings	Yes	Yes	No	Yes
Ability to adapt to unanticipated baseline adjustments	Yes, revise savings calculation procedure	Yes, if changes don't affect measure performance	Can be complex	Yes, most situations can be readily simulated
Measure type	All	Could be difficult for shell measures	All, as long as impact > 10% of total consumption	All
Ease of method understanding by non-technical reviewers	Easy to understand	Easy to understand, as long as baseline adjustments are straight-forward	Easy to understand, as long as baseline adjustments are straight-forward	Difficult
Special skills required	General science or engineering	Metering and monitoring expertise	Statistics or math background required.	Many - including simulation application, metering and monitoring, statistics.
Level of measure interaction	None, unless part of the engineering equation	None	Full interactions present in billing data accounted for	Full interactions present in model accounted for

Table 7.1: Continued

Consideration	Option A	Option B	Option C	Option D
Length of post-retrofit period	Representative period, depending on measured parameter	Representative period	At least one year	None, if model calibrated to pre-retrofit data
Applicable to new construction project	Yes	No	No	Yes
Accuracy	Low, subject to uncertainty calculation	Can be very good, subject to uncertainty calculation	Can be very good, subject to uncertainty calculation	Moderate to good, subject to modeler ability
Cost	Estimate 1% - 3% of annual measure cost savings	Estimate 3% - 15% of annual measure cost savings	Estimate 1% - 10% of annual measure cost savings	Estimate 3% - 10% of annual measure cost savings

Selection of the M&V option represents a balance between accuracy and cost. Approaches for striking this balance vary in the literature. For example, ASHRAE Guideline 14¹⁴⁴ takes a quantitative approach, where the risk in the uncertainty is calculated from the energy value of the difference between the savings estimates at the upper and lower ends of the confidence interval. Improvements to the M&V approach are introduced iteratively, with the incremental M&V costs compared to the reduction in savings risk. The literature on greenhouse gas trading policy encourages more comprehensive M&V by applying savings discount factors tied to the M&V option.¹⁴⁵ The USEPA Conservation Verification Protocols¹⁴⁶ direct the evaluator to report verified savings at the low end of the confidence interval, thus encouraging more precise estimates.

M&V Plan

A measurement and verification plan should be developed for each site included in the M&V study. These individual site plans should be filed with the final EM&V report at the conclusion of the project. The purpose of the plan is to identify the data needs and analysis procedures prior to collecting the field data. The overall components of the M&V plan are listed below.¹⁴⁷

¹⁴⁴ (* ASHRAE 2002).

¹⁴⁵ A discussion of other approaches is found in “International Greenhouse Gas Trading Programs: A Discussion of Measurement and Accounting Issues.” (Vine et al. 2003).

¹⁴⁶ Savings are reported with 75% confidence at the low end of the confidence interval. See (USEPA 1995a).

¹⁴⁷ The M&V plan outline is taken from the 1999 ASHRAE Handbook, Applications Volume, Chapter 39 (ASHRAE 1999). The IPMVP gives less precise guidance on M&V plan content than what is listed here.

Identify Goals and Objectives

The goals and objectives of the M&V activity at a particular site are stated in the plan and may include the following:

- Verification of measure installation
- Verification of proper operation of measures
- Measurement of specific parameters required for energy savings calculations
- Metering energy consumption and demand for energy savings calculations

Specify Building Characteristics

Building characteristics are listed in the plan to help future users of the data understand the context of the monitored data. The building characteristics description should include:

- General building configuration and envelope characteristics (particularly the energy-related characteristics)
- Building occupant information (number of occupants, occupancy schedule, activities)
- Internal loads (lighting, appliances, plug and process loads)
- Type and quantity of energy-using systems and control setpoints
- Changes in building occupancy or operation during the monitoring period that may affect results

Besides the general building or facility description, a description of the energy conservation measures and their respective projected savings should also be included. A typical form used to describe a building is shown in Table 7.2. This form is shown as an example only, and may be too detailed or too simplistic for any particular M&V plan.

Table 7.2: Building Characteristics Reporting Form

General Information	
Building Description	3-Story Municipal Office Building
Building Type	According to CEC load forecasting conventions
Building Vintage	New construction
Building Location	Davis, California
Building Owner	City of Davis
Occupancy Schedule	6 a.m. to 6 p.m. normal business days
Weather Station	Davis Airport
Architectural Features	
Total Conditioned Area	49,000 net SF [per architectural program] / 75,000 gross SF
Layout	3 floors above grade, plus basement
Floor to floor height	15 ft
Floor to ceiling height	10 ft
Overall Opaque Wall U-value	0.054
Overall Roof U-value	0.039
Window U-Value (with frame)	0.68
Window Shading Coefficient	0.60
Window to wall area ratio	23
Internal Load Densities	
Lighting Power Density – All Areas	1.57 W/sf 0.2 W/sf task lighting in office areas
Plug Load Density – All Areas	1.0 W/sf
Occupant Density – All Areas	175 people

Table 7.2: Continued

HVAC Description	
HVAC System Type	Two Packaged VAV with Gas Heating Sections and Return Fans
Fan Flow Control	Inlet vanes
Terminal Heating	Fan Powered boxes w/ Electric Coils
Cooling/Heating Setpoint	74° F Cooling / 72°F Heating
Supply Air Rate	40,000 cfm per unit
Outside Air Ventilation Rate	2,625 cfm (15 cfm per person)
Control Sequence	Night Setback/Economizer
Cooling System	9.47 EER DX Coils
Measure Description	
DEER 2003 Measure ID Number(s)	D03-002 lighting power reduction D03-025 high performance glass and daylighting controls
Measure capacity and efficiency information	D03-002 – 40% LPD reduction D03-025 – Standard glass and no daylighting controls
Operating hours	Follows building occupancy
Installation date	New construction
Baseline capacity and efficiency information	Title 24 requirements by space type

The building description data should be tailored to the scope and needs of the project. It may not be desirable to gather all of the information about the building during a phone survey confirming measure installation. However, future use of the data may depend on a reasonable understanding of the context of the equipment and related operating conditions. Descriptive data should be coordinated with the needs of related “overarching” studies, such as updates to the Database on Energy Efficiency Resources (DEER)¹⁴⁸ and measure potential studies, as described in Chapter 15, Overarching Evaluation Studies.

Specify Data Products and Project Output

The end products of the M&V activity should be specified. These data products should be referenced to the goals and objectives on the project, and include a specification of the data formats and engineering units. To maximize future use of M&V project results, the results should be normalized and reported according to a standard format. For example, the format followed in the DEER Update Study¹⁴⁹ is shown in Table 7.3.

¹⁴⁸ (Xenergy et al. 2001).

¹⁴⁹ (Itron Incorporated 2004).

Table 7.3: Example Units for Measure-Specific Savings Reporting from DEER

Measures	Measure group	Impact Units	Scalable	Comments
Lighting	Lighting	kW connected	Yes	Impacts expressed as % reduction
Lighting controls	Lighting	kW controlled	Yes	Impacts expressed as % reduction
Plug loads	Plug loads	kW connected	Yes	Impacts expressed as % reduction
Roof insulation, cool roof	Shell	SF roof	No	
Air curtain	Shell	Each	No	
Glass solar heat gain	Shell	SF glass	% change in SHGC	SHGC is the “solar heat gain coefficient,” a measure of the solar energy transmitted by a particular glazing system
HP glass plus daylighting controls	Shell	SF glass	No	
Chillers	HVAC	ton	Difference in kW/ton	Impacts scalable by the difference in the chiller efficiency, expressed in kW/ton
Chilled and hot water distribution loop temperature control	HVAC	SF building	No	
Chilled and hot water distribution loop flow control	HVAC	Hp pump	No	
HVAC system conversion	HVAC	SF building	No	
VAV fan control	HVAC	Hp fan	No	
Evap cooling of makeup air	HVAC	Ton	No	
Ventilation rate	HVAC	SF building	No	
Heat recovery	HVAC	Each	No	
Economizers	HVAC	Ton	No	
HVAC maintenance	HVAC	Ton	No	
Heat rejection system	HVAC	Ton	No	
High efficiency furnace	HVAC	KBtu/hr	Difference in AFUE	Impacts scalable by the change in furnace efficiency, express as the Annual Fuel Utilization Efficiency (AFUE)
Large boilers	HVAC	SF building	Difference in efficiency	Impacts scalable by the change in boiler thermal efficiency, express as a percent.
HVAC system control	HVAC	SF building	No	
High efficiency packaged HVAC	HVAC	Ton	Difference in EER	Impacts scalable by the change in cooling efficiency, express as the Energy Efficiency Ratio (EER)
High efficiency motors	Motors	Hp motor	Difference in efficiency	Impacts scalable by the change in motor efficiency, express as a percent
Water heater tank insulation	Water heating	SF building	No	
Water heater efficiency	Water heating	SF building	Difference in efficiency	Impacts scalable by the change in water heater efficiency or energy factor (EF), express as a percent.

Table 7.3 represents an example of the measure impact normalization scheme designed into the DEER study at the time of publication of the Framework. Be sure to consult the latest version of the DEER for the current normalization scheme before planning an evaluation project.

Specify M&V Option

The M&V option chosen for the project should be specified according to the options described in this Chapter:

- Option A - Partially Measured Retrofit Isolation
- Option B - Retrofit Isolation
- Option C - Whole Facility
- Option D - Calibrated Simulation

For any particular building, a combination of options may be used. The M&V option applied to each measure should be described.

Specify Data Analysis Procedures and Algorithms

This is a key component of the M&V plan. Often, data are collected without a clear understanding of the later use for the data. This can result in either extraneous data collection and/or missing data during the data analysis step. Fully specifying the data analysis procedures will help ensure that an efficient and comprehensive M&V plan is presented.

Specify Field Monitoring Data Points

The actual field measurements made are specified here. This is analogous to an energy management system “points list,” where the sensor, location, measurement and engineering units are specified. For example:

- For measuring the run-time of a boiler, the field data point description would be: *“Accumulated run-time of draft fan serving boiler number 1, using an inductive run-time logger mounted on the draft fan motor.”*
- For measuring air conditioner supply air temperature, the field data point description would be: *Duct air temperature (in degrees F) using a sheathed thermistor sensor located in the supply duct three feet downstream from AC-1.*
- For measuring chilled water temperature, the field data point description would be: *“Chilled water supply temperature measured with a probe-type thermistor inserted in a thermowell.”*

Estimate Data Product Accuracy

All measurement systems have error, expressed in terms of the accuracy of the sensor and the recording device. The combined errors should be estimated using a propagation of error analysis, and the final data product accuracy described. For example, a thermistor sensor and recording device may have an accuracy of $\pm 0.5^{\circ}\text{C}$ ($\pm 0.9^{\circ}\text{F}$). If the data product delivered is heat supplied by a boiler, then the combined error of two sensors and the flow measurement device on the overall calculation should be supplied. In situations where the data product is calculated from a temperature difference, the temperature measurement error can represent a large fraction of the error in the final data product. All measurement device accuracy specifications assume the device has been recently calibrated and the sensor is placed correctly.

Specify Verification and Quality Assurance Procedures

Data analysis procedures should include checks to identify invalid data. For example, temperature differences that are negative can indicate a problem with a sensor or with sensor placement. Energy data that exceed the product of the nameplate rating and the monitoring duration are also invalid. Any data verification and quality assurance procedures planned for the project should be described.

Specify Recording and Data Exchange Formats

Data formats should be described, so others attempting to use the data can interpret them correctly. For example, monitoring equipment may store data in a proprietary format, requiring the use of proprietary software to read and analyze the data. The use of non-proprietary data and software is recommended. Where proprietary data is collected, simple solutions often exist for making the data available to the public: e.g., deleting the name and/or address of a building. Some monitoring systems allow export to ASCII or Microsoft Excel formats. These should be specified as part of the monitoring plan.

Measure Installation Verification

All M&V options generally include measure installation verification as a component of the overall M&V plan. The objectives of measure installation verification are to confirm that: (1) the measures were actually installed, (2) the installation meets reasonable quality standards, and (3) the measures are operating correctly and have the potential to generate the predicted savings. Verification activities are generally conducted during on-site surveys of a sample of projects. Phone and mail surveys may be used for very simple measures (such as CFL replacements), but on-site inspection is preferred.

Option A

Savings are determined by partial field measurement of the energy use of the system(s) to which an ECM was applied separate from the energy use of the rest of the facility. Measurements may be either short-term or continuous. Partial measurement means that

some but not all parameter(s) affecting the building's energy use may be stipulated, if the total impact of possible stipulation error(s) is not significant to the resultant savings. Careful review of ECM design and installation will ensure that stipulated values fairly represent the probable actual value. Stipulations should be shown in the M&V Plan along with analysis of the significance of the error they may introduce. Savings are estimated from engineering calculations using short-term or continuous post-retrofit measurements and stipulations. A typical application of Option A is a lighting retrofit, where pre/post fixture watts are stipulated from a standard fixture wattage table, and operating hours are derived from short-term measurements of fixture run-time.

The overall Option A M&V path is shown in Figure 7.2.

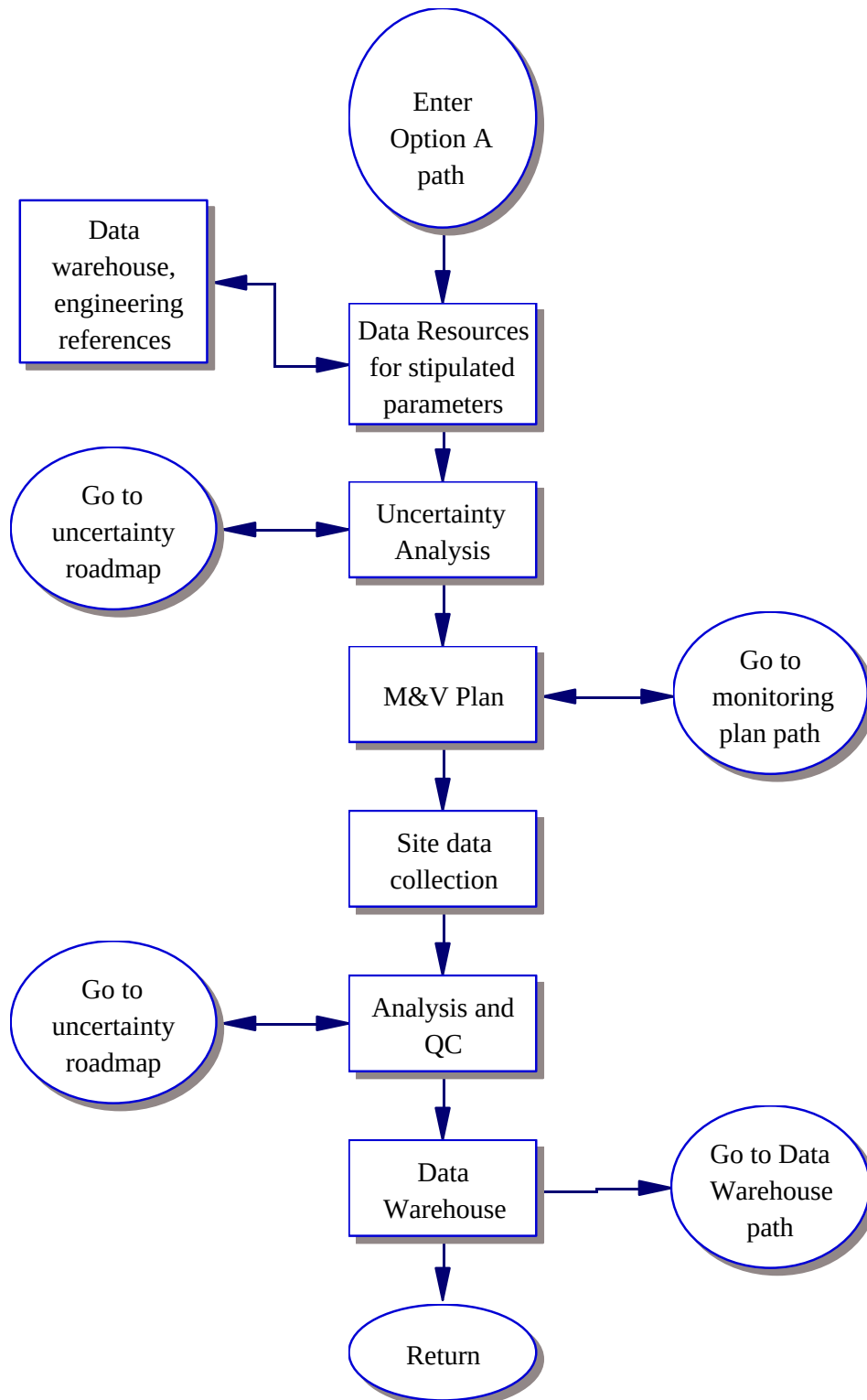


Figure 7.2: Option A Roadmap

The important issue relative to the application of Option A is that **some** measurement must occur and **some** parameter is stipulated. The number of parameters stipulated can vary significantly. For example, measurement of only one parameter with all other

parameters stipulated, or measurement of all parameters but one, with the remaining parameter stipulated are all covered under Option A. When M&V activities are conducted after measure installation, then Option A may be the only option available, since some aspect of the pre-retrofit condition may need to be stipulated.

The best applications for Option A include:

- Measures with constant loads
- Measures with small anticipated impact overall (low risk measure)
- Measures with small anticipated impact relative to the energy recorded at the billing meter
- Measures where interactive effects are small and can be ignored¹⁵⁰
- Measures where baseline adjustments to whole-building data could be problematic
- Studies where uncertainty in the deemed parameters is acceptable

All measures can be verified using Option A, but the accuracy of this option is generally inversely proportional to the complexity of the measure. Typical measures analyzed using Option A include:

- Lighting efficiency
- Lighting controls
- Constant load motors
- HVAC efficiency improvements

Some typical measured and stipulated parameters for measures are shown below.

Table 7.4: Stipulated and Measured Parameters for Lighting Measures

Engineering Parameter	Stipulated data source	Measured data source
Fixture counts	Contractor records	Pre-retrofit field audit Post-retrofit field audit
Fixture wattage	Standard fixture wattage tables	Spot measurements of fixtures or groups of fixtures
Operating hours	Standard values by building type Customer questionnaire Assume measured pre equals post or vice versa	Short-term monitoring of fixture on/off status Short or long-term monitoring of fixture circuits
Interactive factors	Ignored or stipulated based on engineering analysis	Not practical in most cases

¹⁵⁰ Interactive effects include the interactions between the measure and a non-measure-related end use (e.g. efficient lighting generally reduces cooling loads) and interactions between packages of measures that can cause the sum of the measure package savings to be less than the sum individual measure savings.

Table 7.5: Stipulated and Measured Parameters for Motor Measures

Engineering Parameter	Stipulated data source	Measured data source
Motor count	Contractor records	Pre retrofit field audit Post retrofit field audit
Motor wattage	Based on nameplate hp, efficiency and oversizing assumption	Spot motor kW measurement
Operating hours	Standard values by building type, Customer questionnaire, Assume measured pre equals post or vice versa	Short-term monitoring of motor run hours Short or long-term monitoring of motor circuit current or kW
Interactive factors	Ignored or stipulated based on engineering analysis	Not practical in most cases

Table 7.6: Stipulated and Measured Parameters for HVAC Measures

Engineering Parameter	Stipulated data source	Measured data source
Unit count and capacity	Contractor records	Pre-retrofit field audit Post-retrofit field audit
Unit efficiency	Based on nameplate EER or SEER	In-situ efficiency measurement
Operating full load hours	Standard values by building type Assume measured pre equals post or vice versa	Short or long-term monitoring of unit current or kW

Data Resources

Assumptions for stipulated parameters used in Option A come from a variety of sources, including:

- Evaluation data warehouse and/or DEER
- Lighting fixture wattage tables from the Standard Performance Contract program¹⁵¹
- Express Efficiency workpapers¹⁵²
- Engineering references, such as the ASHRAE Handbook
- Manufacturers' data
- General evaluation literature

Data resources used to justify stipulated value assumptions should be documented as part of the M&V plan.

¹⁵¹ The Standard Performance Contract (SPC) program has published a table of fixture watts for standard fixture types. (Pacific Gas & Electric 2001).

¹⁵² Workpapers filed in support of the measure savings assumptions for the Statewide Express Efficiency commercial retrofit program are another source of engineering data and measure performance assumptions for deemed savings estimates. See www.cpuc.ca.gov.

Uncertainty Analysis

The level of uncertainty in savings estimates calculated under Option A is a function of (1) the variance in the deemed values relative to the actual measure performance parameters, (2) the magnitude of any measure or end use interactive effects, (3) the uncertainty in the measured parameters, and (4) any errors introduced when a sample of measures is used. The M&V plan should account for the potential errors in the analysis and develop a strategy to reduce the most influential errors.

Initial estimates of engineering parameter uncertainties should be used to provide an estimate of the overall uncertainty in the savings calculations. Initial estimates of parameter uncertainty values are typically based on professional judgment, due to a general lack of data on the uncertainty of stipulated engineering parameters.¹⁵³ The uncertainty estimate should address both instrument error and variations due to differences in equipment schedule and performance. The uncertainty of the savings estimate will likely be reduced by site-specific data collection. An upfront estimate of the uncertainty in the savings calculations should be used to guide the development of the M&V plan.

Option A Example

This example describes an uncertainty analysis applied to a lighting retrofit project. The annual energy savings from a commercial lighting retrofit program are estimated using an engineering analysis. Initial estimates of the uncertainty of each input parameter are shown in Table 7.7 below.

Table 7.7: Data for Lighting Retrofit Program

Parameter ¹⁵⁴	Value	Uncertainty	Error
Number of fixtures replaced	10,000	± 1%	± 100
Efficient fixture watts	105W	± 1%	± 1.05
Baseline fixture watts	160W	± 7%	± 11.2
Lighting operating hours	2,860 hr	± 25%	± 715
Demand diversity factor of efficient fixture	0.8	± 5%	± 0.04
Demand diversity factor of baseline fixture	0.75	± 5%	± 0.038
HVAC interaction factor	0.24	± 15%	± 0.036
Energy savings	1.28 GWh	± 40%	± 0.51

See the Appendix to this chapter for the calculations associated with this example. The uncertainty in the engineering estimate of savings prior to any M&V activity is ±40%.

¹⁵³ Developing a data resource for engineering parameter uncertainty could be an important role for the Data Warehouse, as described later in this chapter.

¹⁵⁴ See the Appendix to this Chapter for a definition of these parameters.

The last column in the previous table is an indication of the relative influence of each parameter on the overall uncertainty of the estimate. The priorities for data gathering for estimating annual energy savings are presented in Table 7.8.

Table 7.8: Data Gathering Priority Ranking for Gross Annual Energy Savings Estimates

Rank	Parameter	Parameter Uncertainty	Contribution to Overall Uncertainty ¹⁵⁵
1.	Operating hours	± 25%	39%
2.	Baseline fixture watts	± 7%	34%
3.	Demand diversity factor of baseline fixture	± 5%	17%
4.	Demand diversity factor of efficient fixture	± 5%	9%
5.	Efficient fixture watts	± 1%	< 1%
6.	Number of fixtures replaced	± 1%	< 1%
7.	HVAC interaction factor	± 15%	< 1%

To reduce the uncertainty in the overall estimate, data gathering activities should focus on reducing uncertainty in the parameters that have the greatest contribution to the overall uncertainty. The priorities for data gathering for reducing uncertainty are monitoring fixture full load hours, connected load, and demand diversity factors.

Option B

Savings are determined by field measurement of the energy use of the systems to which the ECM was applied; this is separate from the energy use of the rest of the facility. Short-term or continuous measurements are taken throughout the post-retrofit period. Savings are estimated from engineering calculations using parameters derived from the short-term or continuous measurements. A typical application of Option B is a variable frequency drive applied to a constant speed pumping application. Pre-retrofit power consumption is measured with a hand-held power meter (or short-term metering to confirm constant power draw), and post-retrofit power consumption is short-term metered along with some relevant independent quantity (such as fluid or ambient temperature). The relationship between power and the independent variable is used to project long-term post-retrofit energy consumption from the short-term measurements.

The overall path through Option B is shown in Figure 7.3.

¹⁵⁵ See the Appendix to this Chapter for a calculation of the contribution of each parameter to the overall uncertainty.

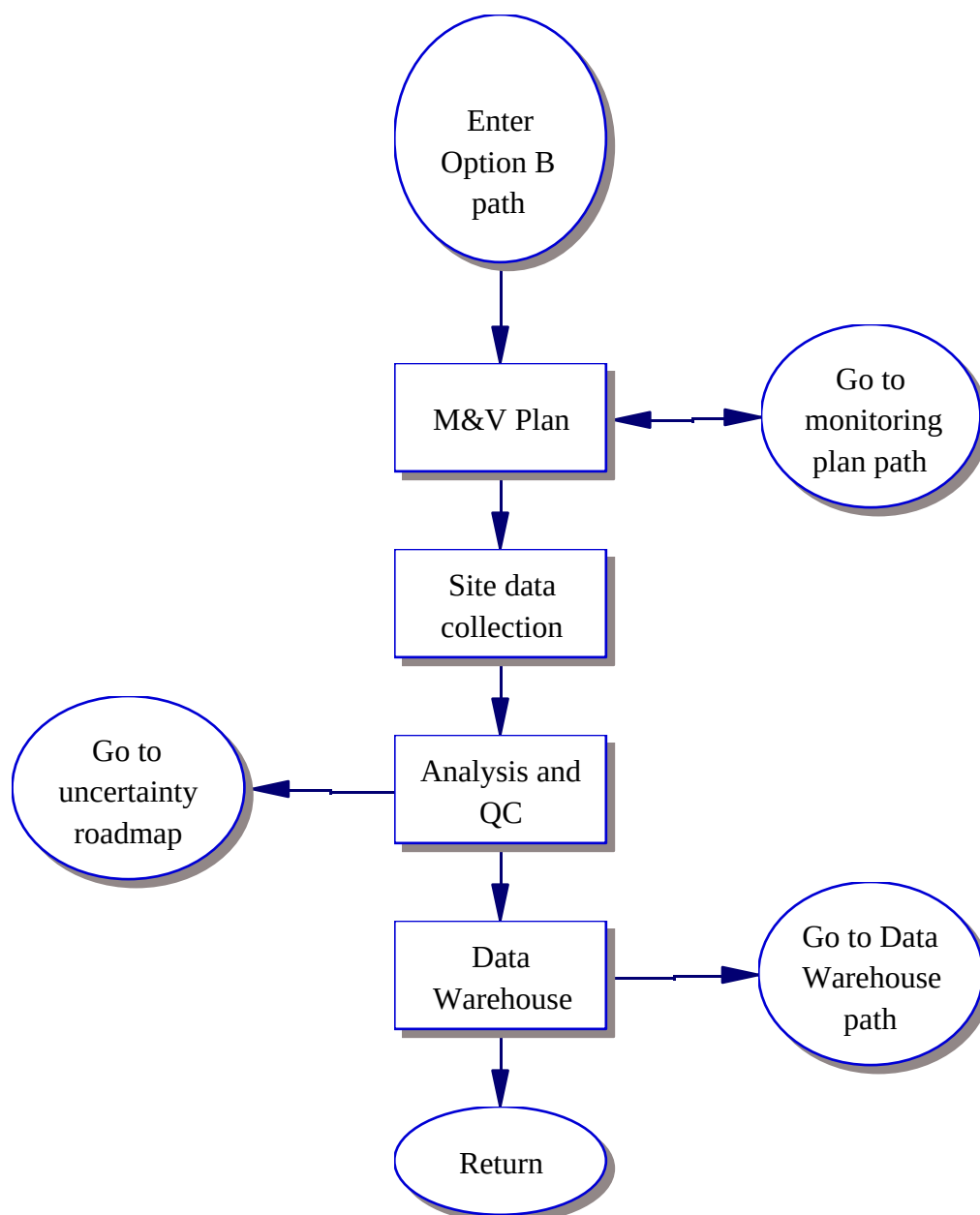


Figure 7.3: Option B Roadmap

Option B is very similar to Option A, but no stipulations are allowed. Full measurement is required. This option is analogous to the retrofit isolation approach described in ASHRAE Guideline 14, where full measurement is also required. The Option B measurements allow true electric power or proxy measurements such as accumulated run-time. Although both of these strategies are allowed within Option B, both may not yield the same results in terms of accuracy.

Short- or long-term monitoring may be employed under Option B. Short-term or spot measurements are appropriate for measures with constant or predictable operating hours. Continuous metering may be installed at sites financed through a performance contract.

Long-term metering provides greater certainty in reported savings, but is more expensive. Secondary use of long-term or continuous measurements may also be used to enhance or maintain measure performance. The benefits of enhanced measure operation may offset the costs of additional metering.

All measures may be verified using Option B, but the degree of difficulty and cost goes up with measure complexity. Savings verification under Option B is likely to be more accurate than Option A, but more costly.

The best applications for Option B include:

- Measures with small anticipated impact relative to the energy recorded at the billing meter
- Measures where interactive effects are small and can be ignored
- Measures where baseline adjustments to whole-building data could be problematic
- Buildings where sub meters already exist to isolate the energy use of affected systems
- Situations where metering added under Option B would have additional benefit to the building operators, offsetting the cost
- Projects where measure level impact information is desired

Option B Example

This example describes an application of Option B to a comprehensive HVAC system retrofit. A chiller plant improvement project at a municipal office building was subject to short-term pre/post monitoring. The retrofit consisted of chiller replacement, cooling tower replacement, conversion of constant volume air handlers to variable air volume (VAV) operation, and conversion of constant volume pumping to variable volume operation. The monitored data were used to build a simple regression model, where daily cooling energy consumption is predicted as a function of average daily ambient temperature, as shown in Figure 7.4. The annual savings were estimated by applying the pre/post regression model to long-term average temperature data.

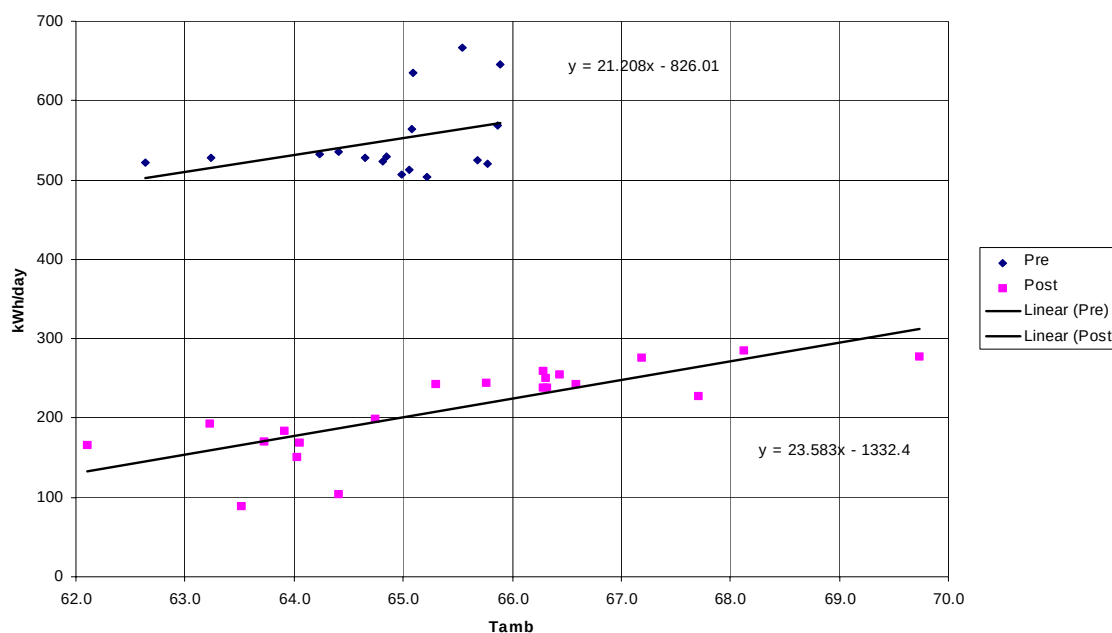


Figure 7.4: Example of Short-Term Monitoring of HVAC System Upgrade Using Option B

Option C

Under Option C, savings are determined by measuring energy use at the whole-facility level. Short-term or continuous measurements are taken throughout the post-retrofit period and compared to 12 to 24 months of pre-retrofit data. Savings are estimated from analysis of whole-facility utility meter or sub-meter data using techniques ranging from simple comparison of utility bills to regression analysis. This approach is very close in concept to the statistical billing analysis technique described in Chapter 6, and many of the issues addressed in Chapter 6 also apply to Option C. A billing analysis conducted under Option C may contain baseline adjustment factors that are specific to a particular building addressed under this option, rather than factors applied to a population of buildings.

The overall roadmap through Option C is shown in Figure 7.5.

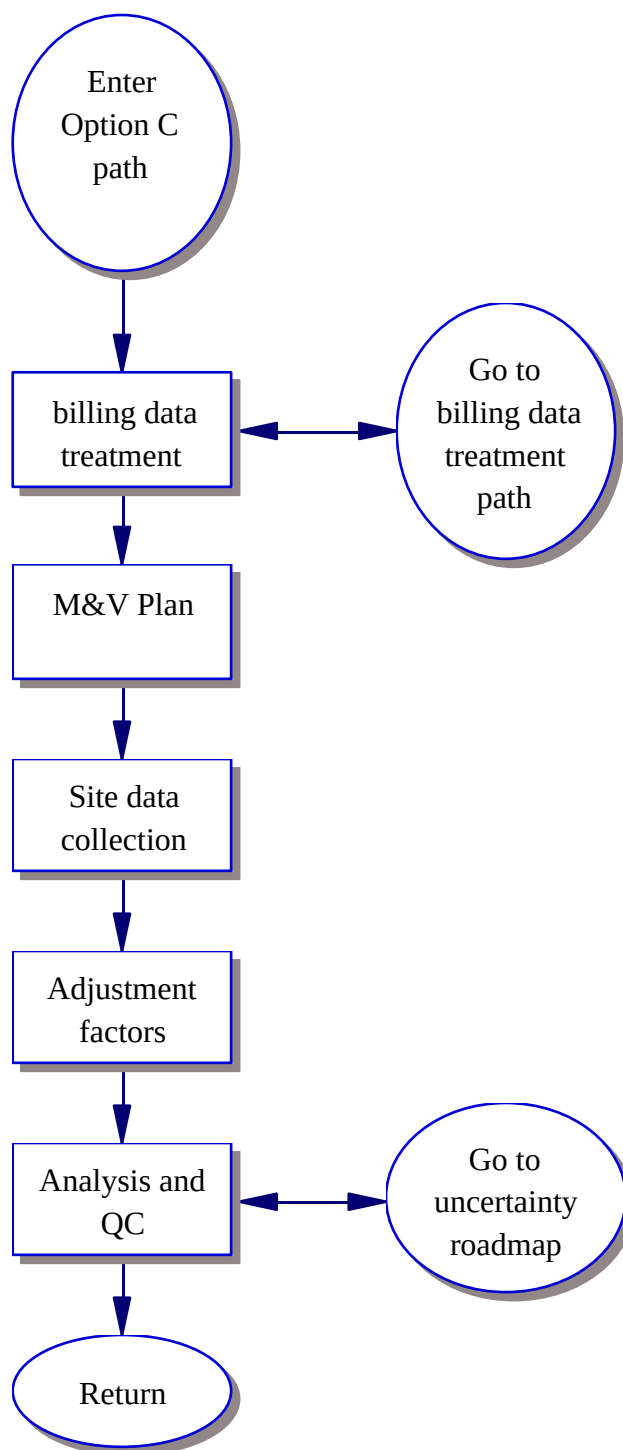


Figure 7.5: Option C Roadmap

Billing Data Collection

Issues with Option C are similar to those encountered when conducting a statistical billing analysis:

- Data may need to be normalized to account for meter read dates
- Missing data or estimated billing may confound the analysis
- Interval demand data from demand-recording meters may be available from the serving utility, but special permission and billing data release permission from the customer or a consultant non-disclosure agreement with the customer's utility will likely be required
- Account information and billing addresses may not match the site studied. During on-site verification activities, recording the meter numbers of all meters affected by the project will help identify the correct billing record

Comparison Models

Simple comparison models look at the monthly billing data (corrected for meter read dates) during the pre- and post-retrofit period, and derive savings as a simple subtraction of the pre- and post-period data. These comparisons are appropriate only for non-weather-dependent measures where the hours of operation and other factors influencing energy consumption besides the energy efficiency measure remain constant during the pre- and post-retrofit period.

Adjustment Factors

Adjustment factors are generally added to the models to account for differences in the pre- and post-retrofit periods that can affect energy consumption outside of the impact of the installed measures. Adjustments are usually made based on weather, hours of occupancy, and building operating mode (e.g. heating or cooling seasonal operating mode). Adjustments for additions of new process loads (such as the addition of a new computer center to an office building), changes in process output (such as widgets produced or hamburgers sold), and occupied floor space may also be required. Weather adjustments may be based on heating and cooling degree days, humidity, and/or temperature.

Analysis Techniques

Simple regression analysis can be used with daily or monthly consumption data. Models developed to predict energy savings from interval data should use daily rather than hourly data to control the number of independent variables and reduce uncertainty. The interval of the weather data used to make the adjustments must be compatible with the billing data; e.g. daily weather data may be needed to calculate weather adjustments to monthly data to correspond to the billing read dates. Monthly models should use pre-retrofit data in full year increments to avoid bias (by capturing all seasonal effects): e.g., 12 or 24

months prior to retrofit. Regression models used for weather correction are usually one of the following types.¹⁵⁶

Two-Parameter

Two-parameter models are used when the energy consumption is linear with outdoor temperature, as shown in equation 1 and Figure 7.6.

$$E = C + B(T)_+ \quad (1)$$

where:

- E = predicted baseline energy consumption per period
- T = outdoor temperature
- C = regression intercept
- B = regression model slope

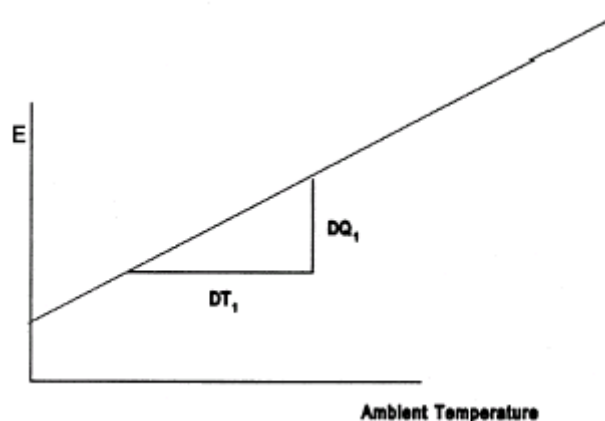


Figure 7.6: Two Parameter Model

Change Point

Change point models are used when the energy consumption data show a constant plus temperature dependent behavior. Electricity consumption in air conditioned buildings with gas heating typically show this behavior, where electricity consumption during the heating season is fairly flat (e.g. not temperature dependent), and shows a temperature dependence in the cooling season. In this situation, a three-parameter change point model is used, as shown in Equation 2 and Figure 7.8.

$$E = C_C + B_C(T - T_{CCP})_+ \quad (2)$$

¹⁵⁶ ASHRAE Guideline 14, Annex D provides detailed information on baseline model development. FORTRAN code to implement several of these models is available from ASHRAE as the Inverse Modeling Toolkit. See “Inverse Modeling Toolkit: Numerical Algorithms,” (Kissock et al. 2003), and “Inverse Modeling Toolkit: Application and Testing,” (Haberl et al. 2003).

where:

E	=	predicted baseline energy consumption per period
C_C	=	constant (non-weather dependant) component of electricity consumption
T	=	outdoor temperature
T_{CCP}	=	cooling change point temperature
B_C	=	slope of the temperature dependent region

Note: the subscript + indicates that only positive values of the quantity in parentheses are used, otherwise the quantity is set to zero.

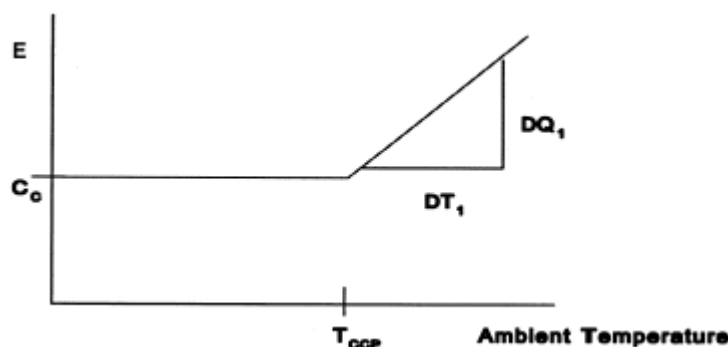


Figure 7.7: Three Parameter Change Point Model

In air conditioned buildings with electric heat, electricity consumption generally varies with temperature in the heating and cooling seasons, and is constant during the “swing” season (periods during the spring and fall when no heating or cooling is needed). In this case, a five-parameter change point model is used, as shown in Equation 3 and Figure 7.8.

$$E = C_{HC} + B_H (T_{HCP} - T)_+ \quad \text{if } T < T_{HCP}$$

$$E = C_{HC} \quad \text{if } T_{HCP} < T < T_{CCP}$$

$$E = C_{HC} + B_C (T - T_{CCP})_+ \quad \text{if } T > T_{CCP} \quad (3)$$

where:

E	=	predicted baseline energy consumption per period
C_{HC}	=	constant (non-weather dependant) component of electricity consumption
T	=	outdoor temperature
T_{HCP}	=	heating change point temperature
B_H	=	slope of the temperature dependent region during the heating season
T_{CCP}	=	cooling change point temperature
B_C	=	slope of the temperature dependent region during the cooling season

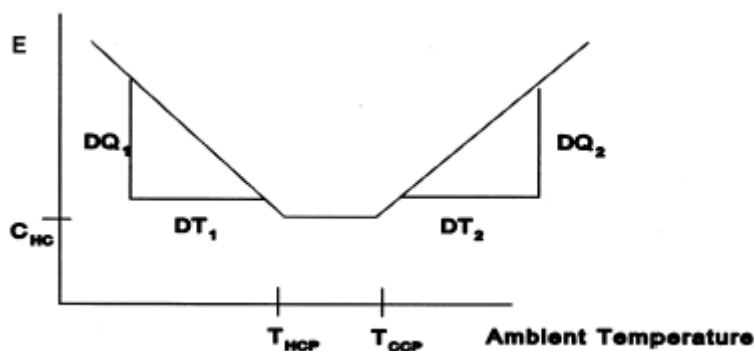


Figure 7.8: Five Parameter Change Point Model

Multi-Parameter

Multi-parameter models are linear regression models that consider weather and other baseline adjustments (e.g., operating hours, production output, etc.). Data on the baseline adjustments included in the model should be collected to determine the magnitude of the adjustments. If the data are not available, the magnitude of the adjustment factor can be stipulated.

Hourly or Sub-Hourly Models

Prediction of baseline hourly or sub-hourly energy consumption may be of interest for peak demand savings estimation. Interval demand meters have been installed in many commercial buildings in California, and these meters represent a potentially rich source of whole-building energy consumption and demand data. Models suitable for predicting hourly or sub-hourly baseline energy consumption include regression¹⁵⁷ and artificial neural networks (ANN).¹⁵⁸ Evaluations of several types of hourly energy prediction models was investigated by ASHRAE under the “Great Energy Predictor Shootout.”¹⁵⁹

Models used to predict hourly whole-building demand may contain terms to account for day of the week, type of day (workday or weekend/holiday), hour of the day, hourly temperature and/or outdoor humidity. Several techniques for predicting baseline hourly demand in order to evaluate the effectiveness of direct load control (DLC) programs¹⁶⁰ were investigated by the California Energy Commission.¹⁶¹ The report examined and tested the ability of several hourly regression models to predict hourly energy consumption in the absence of a DLC action. Hourly models tested included weather correction terms based on outdoor temperature, daily and hourly cooling degree days, and a temperature-humidity index.

¹⁵⁷ A linear regression model scored 2nd place in the Great Energy Predictor Shootout II. See “Modeling Energy Use in Large Commercial Buildings.” (Katipamula 1996).

¹⁵⁸ Examples of ANN models include (Kreider and Wang 1991; Anstett and Kreider 1992).

¹⁵⁹ See the “Great Energy Predictor Shootout.” (Kreider and Haberl 1994)

¹⁶⁰ Although the efficiency programs currently funded through the CPUC specifically exclude DLC programs, the protocols described in this document may be useful if such programs are included in the future.

¹⁶¹ See “Protocol Development for Demand Response Calculation – Findings and Recommendations,” (Kema-Xenergy 2003c).

Model Accuracy Criteria

Regression models developed under Option C are generally evaluated for accuracy according to the three statistical parameters: 1) net determination bias, 2) the root mean squared (RMS) error and 3) the coefficient of variation of the RMS error (CV(RMSE)). These terms are defined below. See also Chapter 12 – Uncertainty for more information.

$$RMSE = \sqrt{\frac{\sum_{period} (M - S)^2}{N}} \quad (4)$$

$$MBE(\%) = \frac{\sum_{period} (M - S)}{\sum_{period} M} \times 100 \quad (5)$$

$$A_{period} = \frac{\sum_{period} M}{N} \quad (6)$$

$$CV(RMSE) = \frac{RMSE_{period}}{A_{period}} \times 100 \quad (7)$$

where:

- M = the measured usage from billing data
- S = the predicted usage from the baseline model
- N = the number of periods in the baseline billing data
- A = the average usage over the period

Baseline Model Selection Criteria

Criteria for developing and selecting baseline energy consumption models under Option C are:

- Extrapolation range - Apply data that are within 90% of the minimum value and 110% of the maximum values used to develop the baseline model.
- Expected savings should exceed 10% of the whole-building energy consumption.
- Baseline period should span at least 12 months, and contain at least 9 data points. Data should be included in full-year increments (e.g. 12, 24, or 36 months) to reduce weather-induced bias.

- The IPMVP specifications for Option C models require the Coefficient of Variation of the Root Mean Squared Error (CV(RMSE)) to be less than or equal to 20% on energy and 30% on demand.

Best applications of Option C include:

- Projects where the expected impacts are 10% or more of the whole-building consumption
- Projects where whole-building rather than measure-specific results are permissible
- Projects where the measures do not lend themselves to retrofit isolation – such as shell measures
- Projects where interactive effects need to be included.

Option D

Savings under Option D are calculated by simulating the energy use of components or the whole facility. Simulation routines should be demonstrated to adequately model actual energy performance measured in the facility. This option usually requires considerable skill in calibrated simulation. Savings are estimated from energy use simulation, calibrated with hourly or monthly utility billing data, and/or end use metering. The overall roadmap through Option D is shown in Figure 7.9.

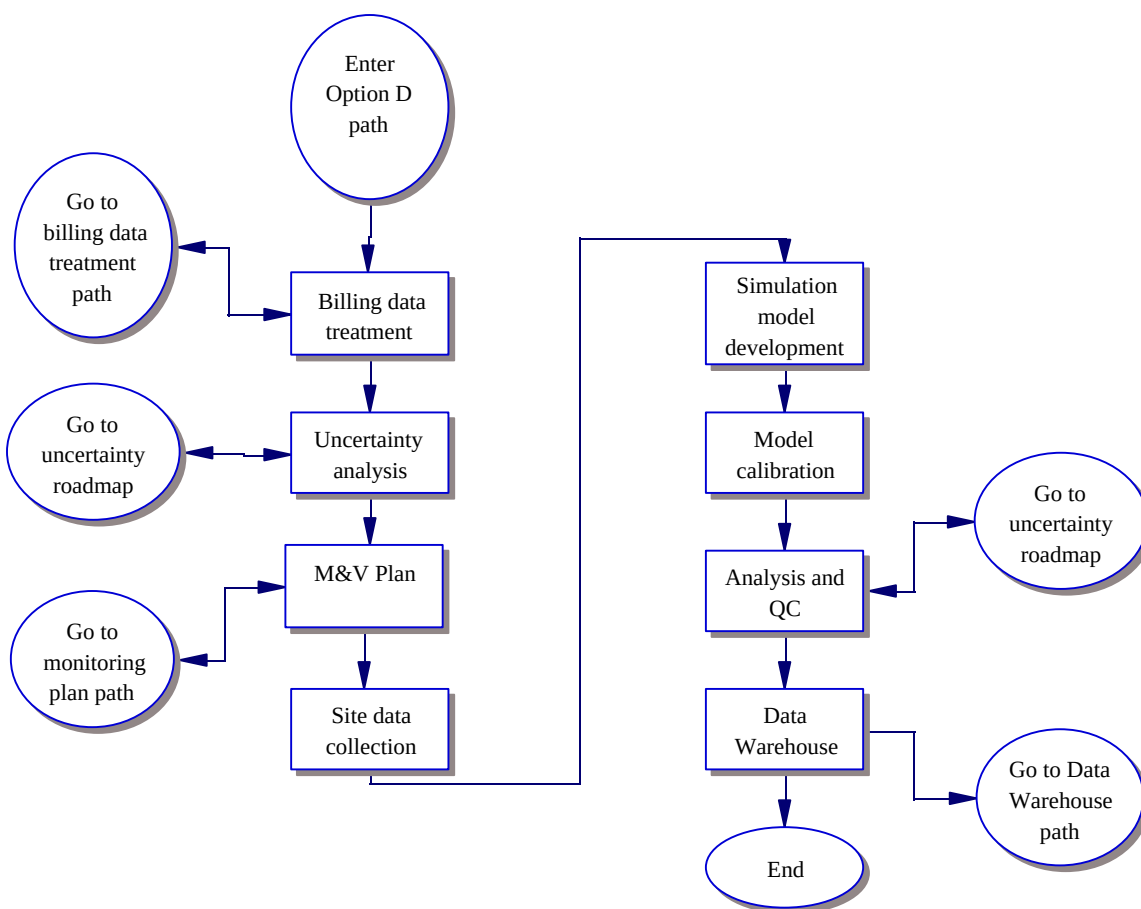


Figure 7.9: Option D Roadmap

Model calibration is done by reducing uncertainty in input variables through measurement and comparing model response to measured data. Certain model inputs are derived from measured data, while other model inputs are adjusted until reasonable agreement to measured data is reached.

There are dozens of model inputs that exert a major influence on the results. Any number of inputs can be adjusted to achieve the same final answer. The problem with model calibration to monthly billing data is that agreement may be reached **for the wrong reason** by randomly adjusting inputs. Accuracy of savings estimates calculated from an improperly calibrated model may be worse than those calculated from an uncalibrated model. The modeler should take advantage of all information available to eliminate as much uncertainty in the model inputs as possible, but should not adjust inputs randomly to produce a predetermined result. Data sources for model calibration are described below.

Billing Data

Historical utility billing data from a one or two year period can be used to check model results for gross errors. Billing data can give some insight into building energy use through a process of elimination, as described below.

- Estimate combined lighting, equipment and fan energy use from billing data for months not requiring mechanical cooling, primarily in systems equipped with economizers and non-electric heating.
- Estimate annual cooling energy from the difference between the non-cooling and cooling month energy consumption.
- Estimate annual heating consumption when heating is the only significant non-electric fuel end use.
- Estimate hot water consumption from billing data during the non-heating season when heating and hot water are served by a non-electric energy source.

Whole Building Hourly Demand Metering

Load research meters installed at larger commercial facilities can provide important calibration data for building energy simulation models. The whole-building load data can also be used to estimate building operating schedules.

One-Time Tests or Spot Metering

One-time tests can be useful to quantify performance of critical building parameters. These tests can be used in conjunction with short-term monitoring to improve model inputs and calibrate model response. They offer a “snapshot” of building performance, but may be useful to infer performance of the building on a long-term basis. One-time tests are best suited to measuring building parameters that do not change with time or change very slowly. Examples of one-time tests include:

Infiltration (blower door) tests. The M&V plan may require a measurement of air leakage rates before and after weatherization actions are taken. Typically, a blower door is used to determine where leaks should be sealed. The blower door is attached to the house, and the leakage rate before any air-sealing actions are taken is established. Buildings already “tight enough” are not subject to any further leak sealing. Leaky buildings undergo sealing until a desirable leakage rate is achieved. Leakage rates before and after leakage sealing are recorded.

Glazing transmissivity tests. It may be difficult to determine the solar gain properties of windows in the absence of building plans and/or specifications. A simple test with a portable pyranometer will provide information on solar transmission, which can be used to estimate the solar heat gain coefficient of the glazing system.

Duct leakage tests. Duct leakage in residential and small commercial buildings can have a significant influence on the overall efficiency of the HVAC system. Duct leakage test methods are described in ASHRAE Standard 152-P.¹⁶²

Refrigerant charge and efficiency tests. The state-of-charge of residential or small commercial air conditioning systems can also affect the efficiency of the unit. Test protocols for evaluating adequacy of refrigerant charge are described in Wilcox et al.¹⁶³

¹⁶² ASHRAE Standard 152-P. (ASHRAE 2001c).

¹⁶³ (Wilcox et al. 2001).

Data collected during refrigerant charge testing and adjustment can be used to estimate the in-situ efficiency.¹⁶⁴

Furnace efficiency tests. Proctor presents a method for determining in-situ furnace efficiency. This test method may prove useful for measuring the effects of furnace replacement programs. The method can be used to establish furnace efficiency for calibrating simulation models.¹⁶⁵

Fan energy tests. One-time tests of fan energy provide valuable data for model calibration. These tests are particularly effective on constant volume fans. Variable volume fan tests will require a manual override of the fan controls to obtain test data over the full range of operation.

Outside air tests. Outside air drawn into an HVAC unit to provide ventilation can have a big effect on total HVAC loads and peak demand. A one-time test of the quantity of outside air may be necessary to reliably establish this model input. Pitot tube traverses in large systems or flow grid measurements¹⁶⁶ in smaller systems can be used.



Figure 7.10: Airflow Measurements Using a Flow Grid

A flow grid is used to measure as-installed airflow rate. A series of flow grids are installed in place of the filters; the airflow rate through each flow grid is displayed on a digital manometer. Flow grids can also be installed in outdoor air intakes to measure outdoor airflow rate.

Short-Term Monitoring

Short-term monitoring of critical performance and behavioral parameters are useful for model calibration. Examples of short-term tests include:

¹⁶⁴ See the web site for the ACRx Service Assistant, www.acrx.com for more information. (Field Diagnostic Services 2003).

¹⁶⁵ “The Development of a Field Furnace Efficiency Test: A More Accurate Prediction of Seasonal Efficiency.” (Proctor 1991).

¹⁶⁶ See “Field Evaluation of a New Device to Measure Air Handler Flow.” (Francisco and Palmiter 2003).

- Status or run-time monitoring of lighting circuits
- Status or run-time monitoring of fans
- HVAC diagnostic tests, including economizer operation, HVAC controls
- Short-term end use monitoring similar to tests conducted under Option B

An example of a short-term diagnostic test of an economizer is shown in the Figure below.

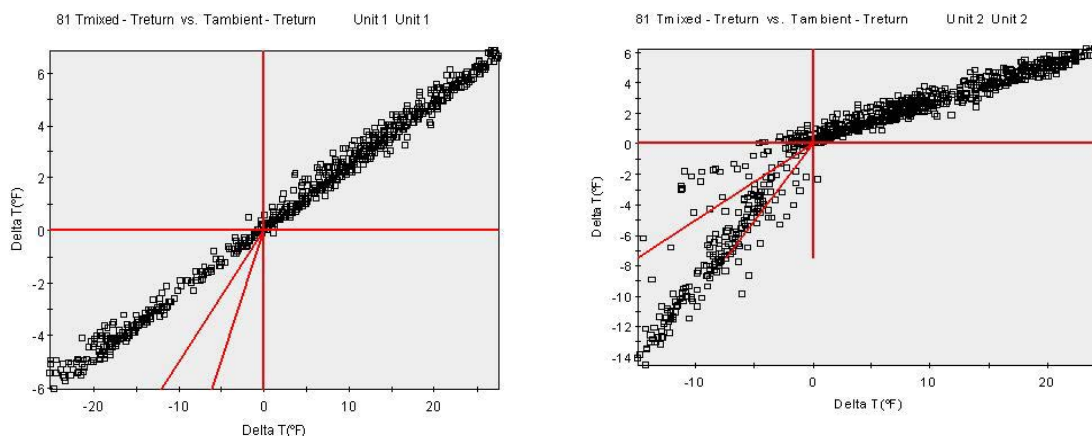


Figure 7.11: Economizer Diagnostic Plots

Short-term monitored data is used to help diagnose system problems. In Figure 7.12, the difference between the cooling coil entering (i.e. mixed) air temperature and the return air temperature ($T_{\text{mixed}} - T_{\text{return}}$) on the vertical (Y) axis is plotted against the difference between the outdoor (ambient) temperature and the return air temperature on the horizontal (X) axis. The slope of the line is equal to the outdoor air fraction. Units with fixed outdoor air (no economizer) have a straight-line relationship between these data, as shown in the plot on the left. Units with economizers show a characteristic change in the slope of the line to the left of the vertical (Y) axis, as shown in the plot on the right. The slope in this region is equal to one, indicating a functioning dry bulb economizer allowing 100% outdoor air.

Calibration Accuracy

Model calibration accuracy is generally assessed by calculating the root mean square (RMS) error and the mean bias error (MBE) between the model output and the calibration data source. These quantities are calculated as follows:

$$\text{RMSE}_{\text{period}} = \sqrt{\frac{\sum_{\text{period}} (M - S)_{\text{hr}}^2}{N_{\text{hr}}}} \quad (8)$$

$$MBE(\%) = \frac{\sum_{\text{period}} (M - S)_{hr}}{\sum_{\text{period}} M_{hr}} \times 100 \quad (9)$$

where:

M_{hr} = the measured hourly subsystem average usage
 S_{hr} = the hourly average predicted usage from the building simulation
 N_{hr} = the number of hours in the monitoring period

See Chapter 12 on Uncertainty for more information on these parameters.

According to ASHRAE Guideline 14, model calibration to monthly whole-building utility data should continue until the RMS error is less than 15% and the net mean bias error is less than 5%. Calibration to hourly data should continue until the RMS error is less than 30% and the net mean bias error is less than 10%.¹⁶⁷

Best applications of Option D include:

- Projects where the expected impacts are greater than the expected modeling error
- Projects where measure-specific results are desired
- Projects where the measures do not lend themselves to retrofit isolation – such as shell measures
- Projects where interactive effects need to be included
- New construction projects, where the baseline must be simulated rather than measured
- Complicated HVAC control measures
- Commissioning and O&M programs

M&V Accuracy Requirements

M&V plans following Options A-D should comply with a minimum set of requirements. An example of the requirements listed in ASHRAE Guideline 14 is shown in Table 7.9.

¹⁶⁷ (* ASHRAE 2002).

**Table 7.9: M&V Accuracy Requirements
(Adapted from ASHRAE Guideline 14)**

Requirement	Option A	Option B	Option C	Option D
Net determination bias		< .005%	< .005%	
Maximum level of uncertainty	±50% at 68% confidence	±50% at 68% confidence	±50% at 68% confidence	±50% at 68% confidence
Baseline model uncertainty			CV(RMSE) < 20% on energy, 30% on demand	
Calibration accuracy				Monthly data: RMS error ± 15% MBE ± 5% Hourly data: RMS error ± 30% MBE ± 10%

Metering and Monitoring

This section on metering and monitoring provides a brief overview of the types of instrumentation used to conduct M&V studies and their application to specific measurement problems. The overall metering and monitoring path is shown in Figure 7.12.

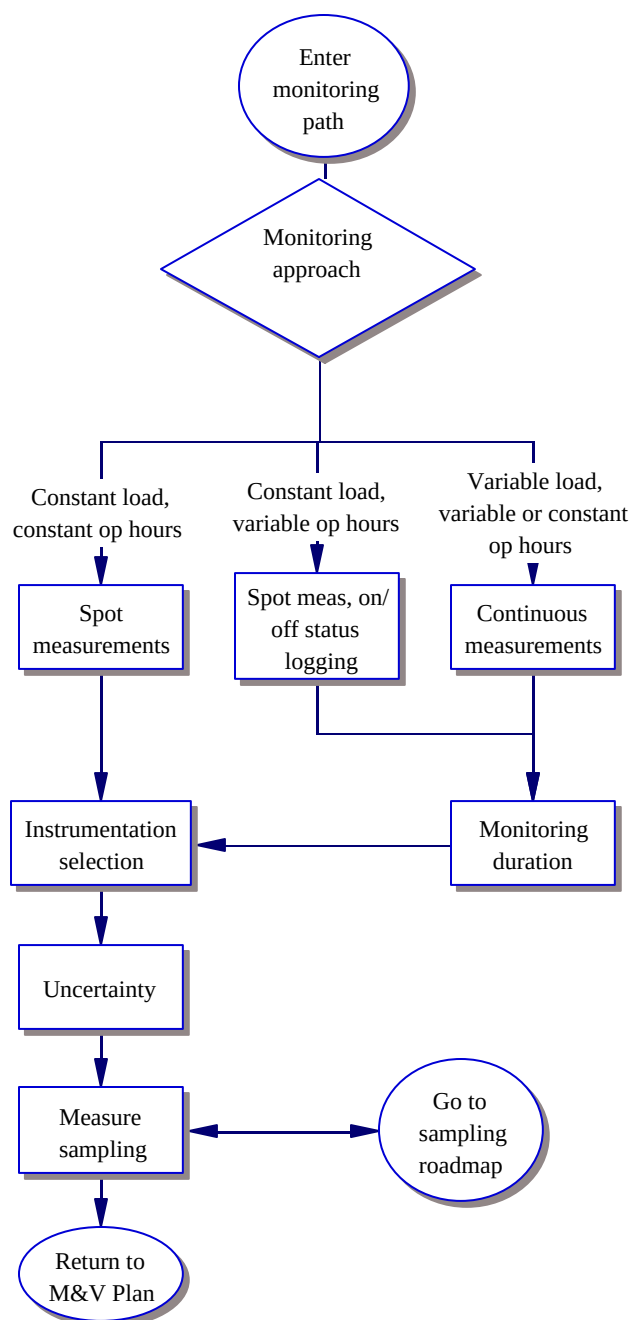


Figure 7.12: Metering and Monitoring Path

Energy efficiency measures can be classified according to their operating characteristics. The characterization helps to determine an appropriate metering and monitoring strategy.

- Constant load, constant operating hours.** Constant load, constant operating hour equipment can be most economically measured with a combination of a one-time measurement of input power (spot watt) combined with a stipulated value for operating hours. Examples of constant load, constant operating hour measures include lighting and motors driving constant volume fans.

- **Constant load, variable operating hours.** Constant load, variable operating hour equipment can be most economically measured with a combination of a one-time measurement of input power (spot watt) combined with on/off status or accumulated run-time recorders. Examples of constant load, variable operating hour measures include lighting on/off controls and motors driving constant volume fans controlled on demand.
- **Variable load, variable operating hours.** Measurement of variable load, variable operating hour equipment generally requires time-series measurements of electric power or a proxy for electric power. One-time tests of electric power and current, combined with time-series measurements of current, can serve as a reasonable proxy for true electric power measurements, as long as the relationship between true electric power and current is fairly constant over the operating range of the equipment. Examples of variable load, variable operating hour equipment include daylight dimming controls, air conditioners and heat pumps, and variable speed drives on pumps and fans.

A summary of energy efficiency measure operating characteristics is shown in Table 7.10.

Table 7.10: Measure Classification by Performance Characteristics

End-Use	Measure Type	Load	Operating Hours	Examples
Lighting	Efficiency	Constant	Constant, Variable	Lamps, ballasts, fixtures
	Control	Constant	Variable	Occupancy sensors, sweeps, time clocks
	Control	Variable	Variable	Daylighting dimming controls
Motors	Efficiency	Constant	Constant, Variable	High-efficiency motors driving constant loads
	Efficiency	Variable	Variable	High-efficiency motors driving variable loads
	Control	Constant	Variable	Time clocks applied to constant load motors
	Control	Variable	Variable	Variable speed drives, time clocks applied to variable load motors
Water Heating	Efficiency	Variable	Variable	Heat pump water heaters
	Control	Constant	Variable	Time clocks, DLC applied to resistance water heaters
	Control	Variable	Variable	Time clocks, DLC applied to heat pump water heaters
	Load Reduction	Constant	Variable	High-efficiency (improved insulation), additional tank wrap, pipe insulation, low-flow fixtures, solar pre-heat applied to resistance water heater
	Load Reduction	Variable	Variable	Tank wrap, pipe insulation, low-flow fixtures applied to heat pump water heater
HVAC	Efficiency	Variable	Variable	High-efficiency compressors, packaged equipment, heat pumps
	Control	Variable	Variable	Economizers, time clocks, reset controls, optimal start/stop
	Load Reduction	Variable	Variable	Insulation, high-performance glass, air leakage sealing applied to buildings with heat pumps and/or air conditioners
	Load Management	Variable	Variable	Thermal energy storage applied to buildings with heat pumps and/or air conditioners
	Control	Constant	Variable	Time clocks, reset controls, optimal start/stop controls applied to buildings with resistance heating only
	Load Reduction	Constant	Variable	Insulation, high-performance glass, air leakage sealing applied to buildings with resistance heating only
	Load Management	Constant	Variable	Thermal energy storage applied to buildings with resistance heating only

Examples

A few examples of measurement strategies are described below. The FEMP M&V Guidelines and ASHRAE Guideline 14 provide extensive examples of monitoring strategies for many energy efficiency measures.

Motor Runtime Logging

The objective of the M&V project is to verify motor loading assumptions and customer-reported run hours as part of an evaluation of a high-efficiency motor program. A sample

of motors in constant-speed, constant load applications is selected for monitoring. Motor running load is measured using a portable watt meter. Magnetically-triggered run-time loggers are placed on the motor for a period of four weeks. Annual run hours are extrapolated from the short-term monitoring, and compared to customer-reported run hours.

Short-Term Lighting Monitoring

Short-term monitoring of building lighting systems is used to verify engineering assumptions of fixture power, demand diversity, and full load hours for a commercial lighting program evaluation. A series of portable, battery-powered data loggers are used to make time-series measurements of lighting circuit current over a two-week period. Circuits are selected at random for monitoring. Spot measurements of circuit kW and current are also made to measure in-situ fixture power and develop an empirical kW per amp conversion factor. This factor is applied to the time-series current data to calculate time-series kW values for each monitored lighting circuit.

Short-Term Monitoring of HVAC Systems

Short-term monitoring of a commercial building HVAC system is used to develop and calibrate a DOE-2 model. A series of portable, battery-powered data loggers are deployed over a four-week period to measure HVAC system parameters, such as temperature, humidity, and HVAC equipment current. Time-series current data are combined with spot kW measurements to calculate time-series kW data. The measured data are used to develop inputs to the DOE-2 simulation model and to calibrate the model response over the monitoring period. The calibrated simulation model is used to estimate the impacts of energy efficiency measures.

Industrial VFD Monitoring

The objective of the project is to develop an empirical model to calculate the energy savings resulting from the application of variable frequency drive (VFD) controls to plastic injection molding machines. A sample of machines is selected, representing a range of machine sizes and operating conditions present in the factory. Input power to each machine is monitored with a portable power monitor for 48 hours before and after installation of the VFD. Since the cycle times of the machines are short (less than one minute), monitoring over numerous cycles is possible during this period. An empirical model is developed to express energy savings in terms of percent idle time. The model is used to estimate energy savings for all VFDs installed at the site.

Monitoring Duration and Schedule

Monitoring duration depends on the measure use variability, climate dependence of the measure, weather variability during the monitoring period, impact estimation precision goals, and budget. Short-term monitoring activities conducted on measures with seasonal variations in use can provide biased data. Monitoring cooling equipment during the cooling season or heating equipment during the heating season are obvious examples. However, short-term monitoring during the hottest part of the cooling season may not

provide enough variation in outdoor temperature to draw meaningful conclusions about annual cooling energy performance.

An example of an experiment conducted on long-term metered data for a lighting system in a commercial building is shown in Figure 7.13.¹⁶⁸ The annual data record was divided into a series of two-week periods, and the results of the two-week data record were extrapolated to annual energy consumption. The deviation from the extrapolated and actual annual consumption for each two-week period is shown in Figure 7.13. In this particular example, errors resulting from data extrapolation from a two-week monitoring period ranged from -22% to 13%. The worst period for two-week short-term monitoring occurred during the December holidays.

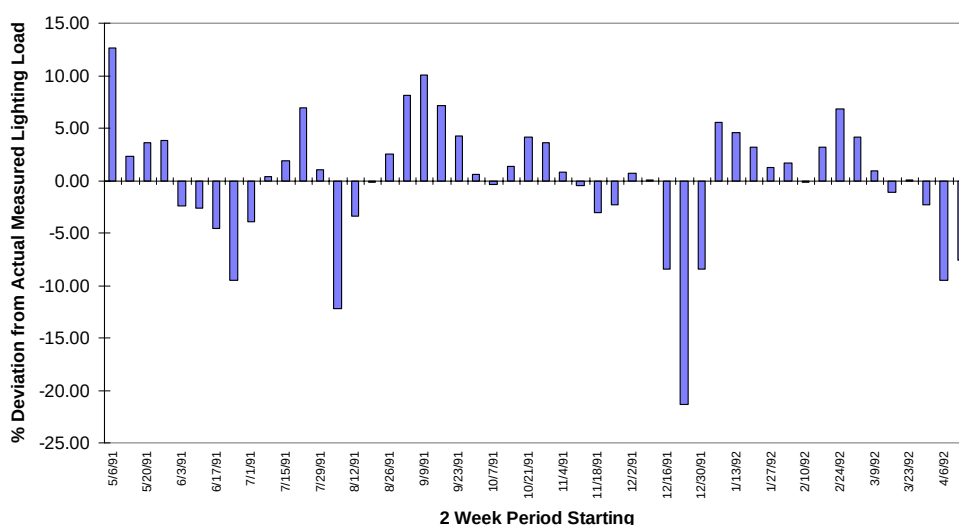


Figure 7.13: Errors Introduced by Short-Term Monitoring

The root mean squared error in the data extrapolation for lighting loads measured in several buildings as a function of the short-term monitoring period is shown in Figure 7.14.

¹⁶⁸ See (Amalfi et al. 1996).

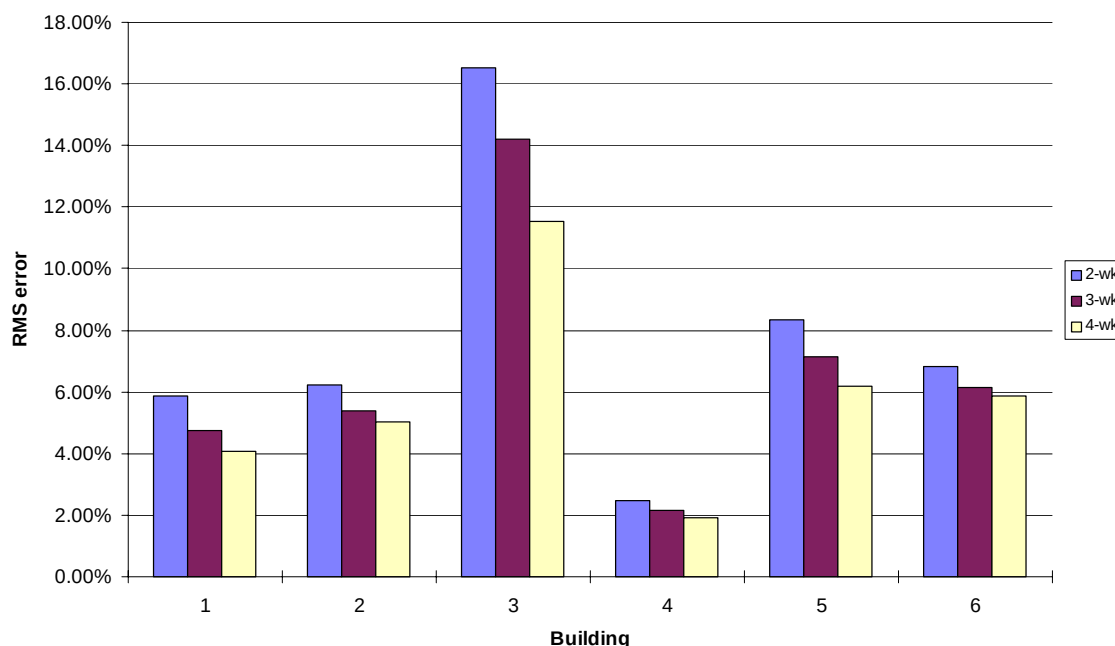


Figure 7.14: Extrapolation Error as a Function of Short-Term Monitoring Period

Increasing the monitoring period beyond two weeks can reduce the root mean squared error (RMSE) on the order of 10 to 30 percent. The majority of the costs involved in performing a short-term lighting study are contained in developing the monitoring plan, installation and retrieval of data loggers, and data analysis. Thus, increasing the study duration (assuming a single deployment of the data loggers) may not have a significant impact on costs, and can help to “smooth out” anomalous short-term behavior.

Instrumentation

A brief summary of instrumentation commonly used in M&V studies is provided below. For more information, consult the FEMP M&V Guidelines,¹⁶⁹ ASHRAE Guideline 14,¹⁷⁰ and Haberl et al.¹⁷¹

Hand-Held Watt Meters

Hand-held watt meters are commonly used to make instantaneous measurements of true electric power. In addition to true electric power, some meters also display voltage, current, kVA, and power factor. These directly measured values convey as-installed, or site-specific information about the equipment. This level of information can be helpful in estimating in-situ performance and reducing the variance between estimated and actual equipment performance due to field installation variances. For example, in lighting retrofit

¹⁶⁹ (* FEMP 2000).

¹⁷⁰ (* ASHRAE 2002).

¹⁷¹ (Haberl et al. 2003).

applications, it is common to find field-measured fixture demands that deviate 30% or more from the rated values available in manufacturers' catalogs. In-situ measurements of motor power often deviate from nameplate values, due to motor over-sizing.

Hand-held instruments are generally used to measure the power of constant load, constant operating hour equipment. When equipment power varies with time, time-series measurements using a recording meter may be necessary. Most hand-held instruments make single-phase power measurements. For three-phase loads, it is often necessary to make measurements on each phase and combine the measurements to obtain the total power. Good quality watt meters provide true power measurements. No adjustments for power factor need to be made to determine the demand because real-time measurement of amps, volts, and power factor are taken simultaneously, providing a true power measurement. Certain types of equipment, such as electronic ballasts, desktop computers, and variable frequency drives create harmonics on the power system. It is important to use a good quality meter capable of RMS power measurements when measuring power on these types of equipment.



Figure 7.15: Hand Held Watt Meter

Watt Transducers

Watt transducers are devices that provide an output signal in proportion to kW and/or kWh. These devices generally do not provide a visual readout, and are designed to be connected to a separate data acquisition or energy management system. Older style analog watt transducers use a Hall-effect device to measure true power. Newer digital watt transducers use high-speed sampling of voltage and current signals to calculate true power. In either case, be sure to examine the accuracy specifications, including the ability to handle harmonic distortion when specifying a watt transducer. Watt transducers can provide an analog DC voltage output, a 4-20 milliamp output, or a pulse output signal.

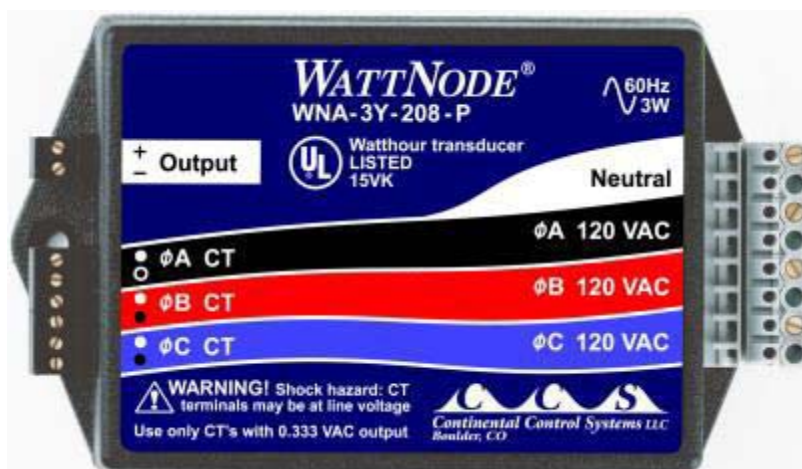


Figure 7.16: Watt Transducer

Data Acquisition Systems

Data acquisition systems are multi-channel instruments that are generally hard-wired into the building electrical system. They are generally wall mounted and connected to various sensors located throughout the building via instrumentation wire. Due to the installation expense, these instruments are used primarily for long-term monitoring projects. They are designed to accept signals from a large variety of sensors, such as temperature, humidity, and flow rate. Most data acquisition systems used in building energy monitoring also have built-in power measurement capability.

Data acquisition system installation requires: (1) mounting the system near the electrical supply panel to obtain system power and reference phase voltages, (2) installing sensors that capture the engineering parameters of interest, (3) connecting the sensors to the system, and (4) calibrating and commissioning the system. Once installed, the system is periodically downloaded by a personal computer over a telephone modem or internet connection. An example of a hard-wired datalogger is shown in Figure 7.17.



Figure 7.17: Hard-Wired Datalogger

Portable Data Loggers

In recent years, small, battery powered data loggers have entered the market. These devices have revolutionized the building measurement industry, providing the capability to make short-term, non-intrusive measurements of building performance at very low cost. Portable data loggers are generally used for **short-term monitoring in temporary installations**. Multiple loggers are generally used, requiring little or no wiring between the logger and the sensors. Loggers are generally left in place for the duration of the monitoring period, then retrieved and downloaded off-site, rather than downloaded remotely.

On/off status and run-time loggers are used in projects involving constant load, variable operating hour devices. These units simply record whether or not a device is “on,” and combined with spot-measurements of input power, allow a calculation of accumulated kWh. The most basic form of these instruments simply record accumulated run-hours. Others record the date and time of each transition from on to off.

Optically triggered run-hour recorders. An optically triggered operating hour recorder is a small device with an optical photocell sensor. The most popular application of the device is monitoring lighting fixture run-hours. To discriminate between fixture and background lighting levels, most units provide an adjustable threshold that will trigger run-time logging whenever lighting levels exceed the threshold value. The device can be directly attached within the fixture or remotely mounted with an optical fiber sensor extender. The basic run-hour logger simply stores accumulated run hours.

Optically triggered on/off recorders. Optically triggered recorders are available which generate a record for every change-of-state. Change-of-state simply means when the light is turned from “off” to “on” or vice versa. Onboard memory stores the time and conditional state into records that can be downloaded. This file can be used to simulate a time-series on/off lighting operating schedule. An example of an optically coupled on/off recorder is shown in Figure 7.18.



Figure 7.18: Optically Coupled On/Off Recorder

Current-triggered recorders. Current-triggered operating hour recorders are available which function similarly to the optically triggered recorders. Both accumulated run-time

and change-of-state recorders are available. Instead of utilizing a photocell to provide the control logic for run-hour accumulation, an internal magnetic field sensor or external current transducer is used. When the device with an internal sensor is placed on an electric motor, it accumulates run-time by sensing the magnetic field from the motor during operation. For devices with external current transducers, the device begins accumulating run-time when a minimum threshold current through the load supply wiring is exceeded. The current transducers are split-core, clothespin type devices attached around the electrical supply wiring, generally at a breaker panel or disconnect box.

Single-channel portable data loggers. Single-channel portable data loggers are small, single-purpose loggers that make time-series measurements of quantities such as temperature, relative humidity, light level, and current. The loggers are called “single-purpose” because they are designed to make only one type of measurement. They can be specified with integral or remote sensors. The loggers are quite small, battery powered, and can be configured for short-term or long-term deployment. The length of the monitoring period is constrained by the size of the memory and the measurement frequency.



Figure 7.19: Single Channel Datalogger

Multi-channel portable data loggers. Small, portable, multi-channel data loggers are available that make time-series measurements of quantities such as temperature, relative humidity, current, and voltage. These devices may be either single- or multi-purpose. Single-purpose loggers are designed with a fixed-channel configuration. Multi-purpose loggers can be configured to make any combination of measurements on each channel. Small multi-channel loggers are available with two to sixteen channels, but four channels is a popular configuration. Some models feature digital real-time displays and sensor excitation power. Monitoring periods are defined by measurement frequency, size of internal memory, and battery capacity. An example of a multi-channel portable datalogger is shown in Figure 7.20.



Figure 7.20: Multi-Channel Portable Datalogger

Measure Sampling

Measure sampling is a key issue in virtually every M&V study. In most cases, it is impractical to monitor every piece of equipment of interest to the study. Some sampling will inevitably be required. Some M&V studies likely to involve measure sampling include:

- Light-logger studies of fixture run-time
- Lighting circuit monitoring
- HVAC monitoring at a facility with multiple packaged air conditioning units.

A sampling plan, as described in Chapter 13, Sampling, should be developed and followed. This will allow quantification of sampling error. Monitoring should not be done only on equipment that is convenient to monitor.

An example of a light logger study conducted at a small commercial building is shown below. A total of twenty-six control points (light switches) were used in the space. Each control point is displayed on the horizontal axis, and the associated connected load (watts of fixture power) are displayed on the vertical axis. A stratified random sampling plan based on the connected watts of each control point was developed. Light loggers were installed in fixtures controlled by the sampled control points as indicated in Figure 7.21.

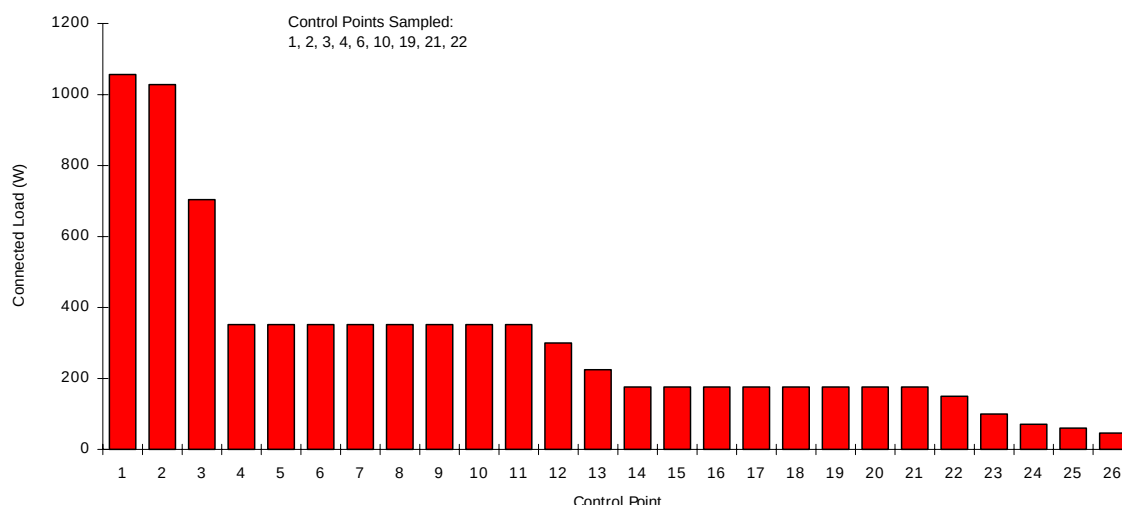


Figure 7.21: Control Point Sampling Strategy for Light Logger Study

Sensor Placement

Sensor placement is a key data acquisition issue. The most accurate sensor and recorder in the industry will not provide good data if the sensor is not placed appropriately. For example, the temperature recorded by a temperature sensor represents an equilibrium condition between the sensor and its surroundings. If a fluid temperature measurement is desired, then the thermal equilibrium temperature of the sensor should be as close to the fluid temperature as possible. In some cases, it may be possible to immerse the sensor in the fluid, or mount the sensor in a thermal well. In other cases, it may be necessary to attach a sensor to a pipe wall, and measure the temperature of the pipe surface as a proxy for the fluid temperature. In this case, a tight connection between the sensor and the pipe surface is required, with sufficient insulation around the sensor and pipe such that the sensor temperature is as close to the fluid temperature as possible. Other examples of sensor placement issues include:

- Air temperature measurements in ducts with non-uniform (stratified) air temperatures. This is especially important in the mixing section (where return air and outdoor air mix upstream of a heating or cooling coil) of an air handler or HVAC unit
- Flow meters installed too close to a flow obstruction
- Current transducers used in electric power measurements facing the wrong way
- Outdoor temperature sensors with inadequate shielding from direct solar radiation, or located in an exhaust air stream
- Optically coupled light loggers mounted where they could be triggered by daylight as well as electric light

Measurement Uncertainty

Total measurement uncertainty should be calculated using a propagation of error technique as described in Chapter 12, Uncertainty. Measurement uncertainty calculations should include the following factors:

- Sensor accuracy
- Recording system accuracy
- Data display or recording resolution
- Sampling error

See ASHRAE Guideline 14, Annex C for additional information on calculating measurement errors.

Quality Control

M&V quality control procedures include activities up and down the data collection and analysis process. The quality control plan should address the following issues as applicable:

- Sensor calibration
- Data quality control procedures, as outlined in the M&V plan
- Baseline model data fit criteria: mean bias error, RMS error, CV(RMSE)
- Checklist on data formatting and reporting requirements

Data Warehouse

A “data warehouse” for compiling and disseminating M&V study results should be established. This center can serve as a resource for the evaluation community, allowing the broad dissemination of evaluation data to improve deemed savings estimates and stipulated parameter assumptions used in M&V, program design, and portfolio planning. Use of quality M&V data resources can reduce the need for redundant data collection and focus resources on issues requiring additional study.

The overall roadmap through the data warehouse path is shown in Figure 7.22.

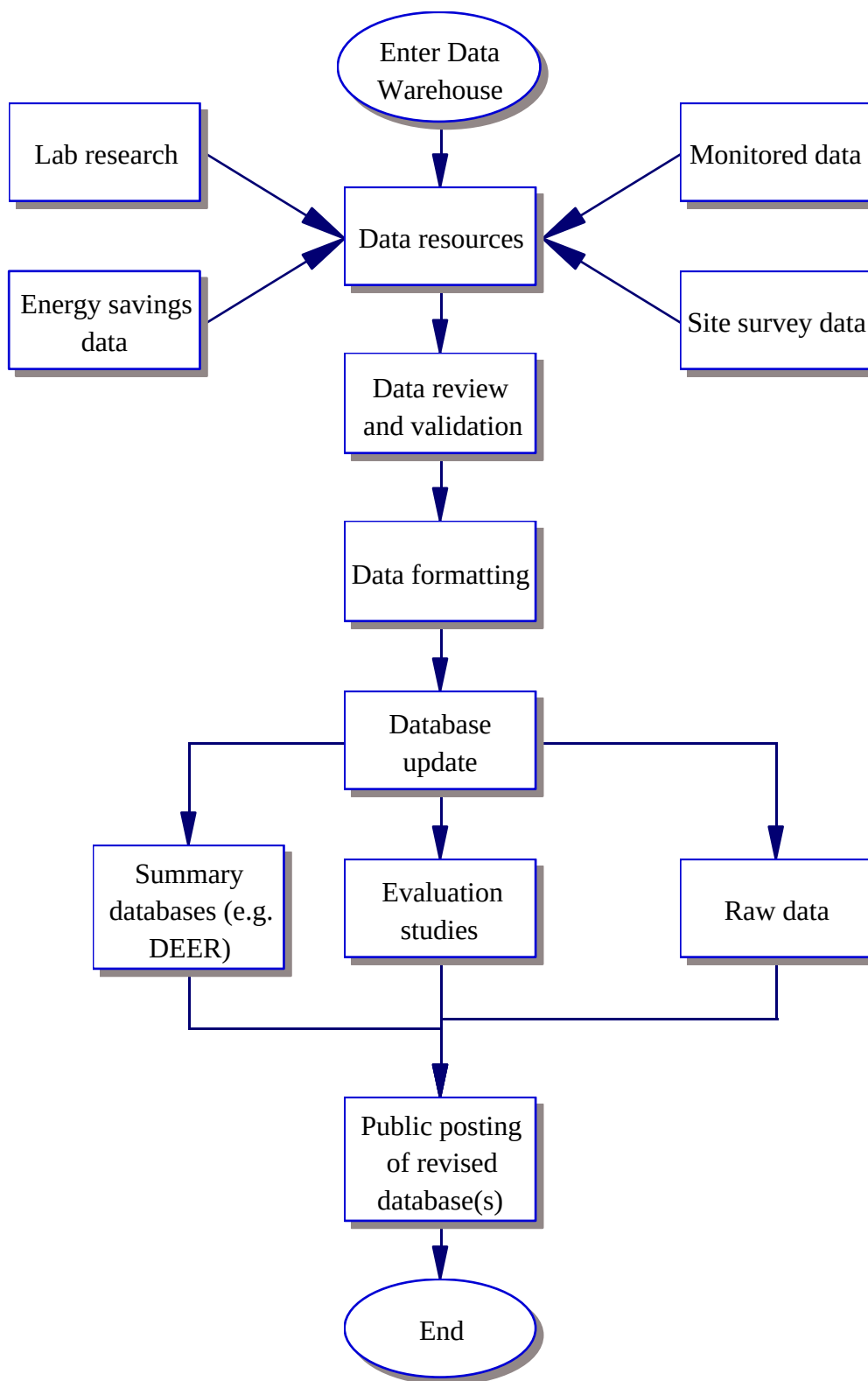


Figure 7.22: Data Warehouse Roadmap

The data warehouse can serve as a clearinghouse for engineering data related to measure performance. These data, or other data submitted and approved in the program implementation plan, can form the basis of the stipulated parameters described in the IPMVP Option A protocols. The relative precision of the parameters and the limits of applicability will form the basis of the uncertainty analysis performed during development of the M&V plan. The data warehouse will also contain data on building performance (total energy consumption) that can be used to examine the link between the presence of measures and the energy consumption of the building.

Data Resources

Data for the data warehouse can come from a variety of sources as appropriate. These are summarized below.

Laboratory Studies

Certain engineering data may be best developed under controlled conditions in a laboratory. For example, the sensitivity of the efficiency of an air conditioner to refrigerant charge and air flow variation may be studied in a laboratory, using instrumentation and test protocols that cannot be easily duplicated in field. Once this relationship is established, it can be applied to field measurements of refrigerant charge and air flow to estimate the impacts of correcting these problems.

Engineering Field Studies

Studies requiring more instrumentation and data analysis than can be reasonably supported under an M&V study can be separately funded and conducted. Examples include performance verification of emerging technologies.

Field Data

M&V data collection to establish engineering parameters for energy savings calculations and unit energy savings form the basis of the data warehouse. Data such as equivalent full load operating hours for HVAC equipment in various climates and applications, energy savings from daylighting controls in specific building types, and the variability of field measured versus nameplate efficiency for measures promoted by efficiency programs will be very useful data for both improving program design and focusing M&V resources appropriately.

On-Site Survey Data

Many program activities involving building energy simulation include whole-building surveys of building characteristics. These data can be a valuable resource for a variety of studies. One example of an existing database of survey information is the Nonresidential New Construction (NRNC) Database, a collection of 990 on-site building surveys covering commercial new construction in California.¹⁷² Development of similar databases of building survey information can be of great value for conducting efficiency potential studies, building codes and standards upgrade analyses, etc.

¹⁷² Available at www.calmac.org. (RLW Analytics 2001).

Data Review, Validation, and Analysis

An authority (such as CALMAC) should be mandated and funded to maintain performance measurement data in the data warehouse. New studies reviewed and approved by the authority can be accepted into the database. This oversight will provide quality control and confidence in the quality of the data. The authority should periodically review gaps in the engineering data, and suggest specific data collection and analytical research projects to fill the gaps to reduce the uncertainty in engineering parameters.

Data Formatting

Unit energy savings and engineering parameters collected during the M&V analysis should include a reference indicating their source, uncertainty estimates when available, and limits of their applicability. These data should be delivered and stored in a standard format. The reporting format suggested in the M&V plan section above is designed to provide sufficient background information to accompany the measure performance data. Unit savings estimates, using the designated units shown in Table 7.2, along with other engineering parameters should be formatted and stored in the database. One promising standard for formatting whole-building survey data used in defining building energy simulation models is the gbXML standard.¹⁷³

Database Update

The authority should update the M&V data database on a periodic basis, corresponding to the submittal of EM&V reports and the conclusion of special research studies. Other databases referencing the EM&V database such as the DEER¹⁷⁴ should be updated in a timely manner to provide feedback to program planning, efficiency potential studies, and subsequent EM&V studies. The data warehouse should also contain a collection of EM&V reports (similar to the current CALMAC web site) and raw M&V data as appropriate.

Data Access

Access to the database should be unlimited for formatted data products such as DEER. Access to more technical data products may be restricted based on user registration and an appropriate disclaimer.

¹⁷³ gbXML was developed by Geopraxis (Petaluma, CA), with funding from the California Energy Commission under the PIER program. (Green Building XML (gbXML) 2003).

¹⁷⁴ (Xenergy et al. 2001).

Appendix to Chapter 7: Uncertainty Calculation Example

This example describes an uncertainty analysis applied to a lighting retrofit project. The annual energy savings from a commercial lighting retrofit program are estimated using an engineering analysis. The gross annual energy savings are calculated from:

$$\Delta kWh = units \times \left[\frac{(Watts \times DF)_{base} - (Watts \times DF)_{ee}}{1000} \right] \times FLH \times (1 + HVAC_c) \quad (10)$$

where:

- ΔkWh = gross annual energy savings
- $units$ = number of units installed under the program
- $Watts_{ee}$ = connected load of energy efficient unit
- $Watts_{base}$ = connected load of baseline unit(s) displaced
- FLH = full load operating hours
- DF = demand diversity factor
- $HVAC_c$ = HVAC system interaction factor for annual energy consumption

Engineer parameter values and their associated uncertainties are listed below.

Table 7.11: Data for Lighting Retrofit Program

Parameter	Value	Uncertainty	Error
$units$	10,000	$\pm 1\%$	± 100
$Watts_{ee}$	105W	$\pm 1\%$	± 1.05
$Watts_{base}$	160W	$\pm 7\%$	± 11.2
FLH	2,860 hr	$\pm 25\%$	± 715
DF_{ee}	0.8	$\pm 5\%$	± 0.04
DF_{base}	0.75	$\pm 5\%$	± 0.038
$HVAC_c$	0.24	$\pm 15\%$	± 0.036

Using the data in Table 7.11:

$$\begin{aligned} \Delta kWh &= 10,000 \times \left[\frac{(160 \times 0.75) - (105 \times 0.80)}{1000} \right] \times 2860 \times (1 + 0.24) \\ &= 1.28 \text{ GWh} \end{aligned}$$

The uncertainty in the estimate is calculated from a propagation of error analysis, as outlined in Chapter 12, Uncertainty. The calculations are reviewed here.

Consider an engineering equation with the general form:

$$R = f(p_1, p_2, p_3 \dots) \quad (11)$$

where:

$$\begin{aligned} R &= \text{result} \\ p_1 &= \text{parameter 1} \\ p_2 &= \text{parameter 2} \\ p_3 &= \text{parameter 3, etc.} \end{aligned}$$

The equation for estimating the error in the overall calculation from the error in the individual parameters is shown below:

$$e_R = \sqrt{\left(\frac{\partial R}{\partial p_1} \times e_{p_1}\right)^2 + \left(\frac{\partial R}{\partial p_2} \times e_{p_2}\right)^2 + \left(\frac{\partial R}{\partial p_3} \times e_{p_3}\right)^2 + \dots} \quad (12)$$

where:

$$\begin{aligned} e_R &= \text{error in results} \\ e_{p1} &= \text{error in parameter 1} \\ e_{p2} &= \text{error in parameter 2, etc.} \end{aligned}$$

The error in the gross annual energy savings calculation is determined from:

$$e_{\Delta kWh} = \left[\left(\frac{\partial \Delta kWh}{\partial \text{units}} \times e_{\text{units}} \right)^2 + \left(\frac{\partial \Delta kWh}{\partial \text{Watts}_{base}} \times e_{\text{Watts}_{base}} \right)^2 + \left(\frac{\partial \Delta kWh}{\partial \text{Watts}_{ee}} \times e_{\text{Watts}_{ee}} \right)^2 + \left(\frac{\partial \Delta kWh}{\partial DF_{base}} \times e_{DF_{base}} \right)^2 + \left(\frac{\partial \Delta kWh}{\partial DF_{ee}} \times e_{DF_{ee}} \right)^2 + \left(\frac{\partial \Delta kWh}{\partial FLH} \times e_{FLH} \right)^2 + \left(\frac{\partial \Delta kWh}{\partial HVAC_c} \times e_{HVAC_c} \right)^2 \right]^{\frac{1}{2}} \quad (13)$$

Using the data in Table 7.11, the partial derivatives are evaluated as follows:

$$\begin{aligned} \frac{\partial \Delta kWh}{\partial \text{units}} &= \left[\frac{(\text{Watts} \times DF)_{base} - (\text{Watts} \times DF)_{ee}}{1000} \right] \times FLH \times (1 + HVAC_c) \\ &= \left[\frac{(160 \times 0.75) - (105 \times 0.80)}{1000} \right] \times 2860 \times (1 + 0.24) \end{aligned}$$

$$= 128$$

$$\frac{\partial \Delta kWh}{\partial Watts_{base}} = units \times DF_{base} / 1000 \times FLH \times (1 + HVAC_c)$$

$$= 10,000 \times 0.75 / 1000 \times 2860 \times (1 + 0.24)$$

$$= 26,600$$

$$\frac{\partial \Delta kWh}{\partial Watts_{ee}} = - units \times DF_{ee} / 1000 \times FLH \times (1 + HVAC_c)$$

$$= -10,000 \times 0.80 / 1000 \times 2860 \times (1 + 0.24)$$

$$= -28,370$$

$$\frac{\partial \Delta kWh}{\partial DF_{base}} = units \times Watts_{base} / 1000 \times FLH \times (1 + HVAC_c)$$

$$= 10,000 \times 160 / 1000 \times 2860 \times (1 + 0.24)$$

$$= 5.67 \times 10^6$$

$$\frac{\partial \Delta kWh}{\partial DF_{ee}} = units \times Watts_{ee} / 1000 \times FLH \times (1 + HVAC_c)$$

$$= 10,000 \times 105 / 1000 \times 2860 \times (1 + 0.24)$$

$$= 3.72 \times 10^6$$

$$\frac{\partial \Delta kWh}{\partial FLH} = units \times \left[\frac{(Watts \times DF)_{base} - (Watts \times DF)_{ee}}{1000} \right] \times (1 + HVAC_c)$$

$$= 10,000 \times \left[\frac{(160 \times 0.75) - (105 \times 0.80)}{1000} \right] \times (1 + 0.24)$$

$$= 446$$

$$\frac{\partial \Delta kWh}{\partial HVAC_c} = units \times \left[\frac{(Watts \times DF)_{base} - (Watts \times DF)_{ee}}{1000} \right] \times FLH$$

$$= 10,000 \times \left[\frac{(160 \times 0.75) - (105 \times 0.80)}{1000} \right] \times 2860$$

$$= 1.029 \times 10^6$$

The individual error terms are calculated and summarized below:

Table 7.12: Summary of Error Calculations

Parameter	$\frac{\partial R}{\partial p_i}$	e_{p_i}	$\left(\frac{\partial R}{\partial p_i} \times e_{p_i} \right)^2$
<i>units</i>	128	± 100	1.63×10^8
<i>Watts_{ee}</i>	-28,370	± 1.05	8.87×10^8
<i>Watts_{base}</i>	26,600	± 11.2	8.88×10^{10}
<i>DF_{ee}</i>	3.72×10^6	± 0.04	2.22×10^{10}
<i>DF_{base}</i>	5.67×10^6	± 0.038	4.53×10^{10}
<i>FLH</i>	446	± 715	1.02×10^{11}
<i>HVAC_c</i>	1.029×10^6	± 0.036	1.37×10^9

The overall error in the annual energy savings estimate is calculated as follows:

$$e_{\Delta kWh} = \sqrt{1.63 \times 10^8 + 8.87 \times 10^8 + 8.88 \times 10^{10} + 2.22 \times 10^{10} + 4.53 \times 10^{10} + 1.02 \times 10^{11} + 1.37 \times 10^9}$$

$= 0.51 \text{ GWh}$ Expressed as a fraction of the estimate:

$$\frac{e_{\Delta kWh}}{\Delta kWh} = \frac{0.51}{1.28}$$

$$= 0.40$$

The uncertainty in the engineering estimate of savings prior to any M&V activity is $\pm 40\%$. The last column in the previous table is an indication of the relative influence of each parameter on the overall uncertainty of the estimate. To reduce the uncertainty in the overall estimate, data gathering activities should focus on reducing uncertainty in the parameters that have the greatest contribution to the overall uncertainty. The priorities for data gathering for estimating annual energy savings are presented in Table 7.13.

**Table 7.13: Data Gathering Priority Ranking for
Gross Annual Energy Savings Estimates**

Rank	Parameter	Parameter Uncertainty	Contribution to Overall Uncertainty
1.	FLH	$\pm 25\%$	39%
2.	$Watts_{base}$	$\pm 7\%$	34%
3.	DF_{base}	$\pm 5\%$	17%
4.	DF_{ee}	$\pm 5\%$	9%
5.	$Watts_{ee}$	$\pm 1\%$	< 1%
6.	$units$	$\pm 1\%$	< 1%
7.	$HVAC_c$	$\pm 15\%$	< 1%

Chapter 8: Process Evaluation

Preface

The process evaluation is an important tool in the evaluation toolbox. The process evaluation consists of in-depth examinations of the design, delivery, and operations of energy programs in order to improve the ability of the program to achieve energy savings and accomplish other program goals. The process evaluation also provides a vehicle for sharing program design and operational improvements with other professionals in the field. When process evaluation results are shared with other energy efficiency professionals, these professionals can assess the relevance of the evaluation findings and recommendations to their policies, programs, and program portfolios. This is especially true for program designers and managers who may want to determine if the evaluation results can be used to improve the cost-effectiveness of their programs.

As with other chapters in this document, the process evaluation chapter is written for a wide range of stakeholders within the California energy program arena. This chapter should be read by regulatory staff and policy makers who need to understand the importance, purpose, scope, and to a limited degree, the tools associated with the process evaluation, and those who are in a position to improve energy programs. This chapter is also intended for evaluation planners and designers as well as the staff who must conduct process evaluations. This is especially true for evaluators who are new to the process evaluation field. Program administrators who may be responsible for funding or coordinating the process evaluation effort would also benefit from reading this chapter. Likewise, program managers who are requested to cooperate with the evaluation effort by providing access to program records and staff during the evaluation process would be helped by being familiar with this chapter. We also suggest that the evaluators of information and educational programs and evaluators who conduct market transformation evaluations read this chapter. These individuals may need to coordinate their evaluation efforts with the process evaluation efforts when similar types of information are needed or when similar evaluation activities are conducted across their evaluation efforts.

This chapter presents a description of the skills needed to conduct process evaluations, a definition of process evaluation, and a discussion of the reasons why process evaluations are a component of the Framework. The chapter then provides a presentation of the typical evaluation tools, types of process evaluations, and examples of evaluation approaches. The chapter ends with a presentation of the recommended decision steps associated with planning and conducting process evaluations within the Framework followed by a discussion on process evaluation budgets.

Skills Required for Process Evaluations

The investigative processes associated with designing, managing and conducting process evaluations focus on a wide range of researchable issues. These issues can range from evaluating the ability of a program's data management system to support the informational needs of the program to assessing if the program is well-designed, managed, targeted, marketed, and operated. As a result, the skills needed to conduct process evaluations are varied, suggesting the need for different skills for different types of process evaluations.

Evaluations that focus on the design and operation of program information systems, for example, need evaluators that understand how information management and information availability influence a program's operations. However, the evaluators should also be skilled at designing, developing, and implementing information systems in order to recommend changes that improve the program's ability to cost-effectively achieve its goals. Process evaluators who assess program satisfaction levels need to have the skills to identify and analyze different program characteristics that influence satisfaction and be able to identify those characteristics that can be changed to improve satisfaction scores. Likewise evaluators that focus on assessing program targeting, marketing and promotional operations need to have skills necessary to assess information flow, content and presentation effects as well as the skills associated with understanding how markets and market segments operate and can be influenced through different outreach and promotional efforts. These examples demonstrate the need to match the skills of the process evaluator with the research goals of the specific process evaluation.

It is equally important that process evaluation managers be trained and/or experienced with the tools used in the process evaluation. For example, if a telephone survey is needed, evaluators need to be knowledgeable and experienced in the field of survey research and instrument design. If focus groups are needed evaluators should be knowledgeable and experienced in the field of focus group design and operation.

Because of the diversity of researchable issues associated with conducting process evaluations and the diversity of skills needed to address these issues it is difficult to define a specific set of skills needed to conduct these evaluations. Instead, the Framework recognizes that a diverse set of program assessment and information analysis skills are needed across the various investigative issues on which these evaluations typically focus. However, in general, the process evaluator should have the following knowledge and skills:

- Expert knowledge of a wide range of energy efficiency programs and a strong understanding of their operational designs, management practices and program goals,
- Expert knowledge of different process evaluation data collection methods and approaches, and a working knowledge of the process evaluation literature and how evaluation approaches have been applied in the energy efficiency program field,

- Strong analysis capabilities and an expert understanding of cause and effect relationships that impact the ability of energy efficiency programs to cost-effectively accomplish their goals,
- A strong understanding of statistical analysis approaches and analytical procedures appropriate for the process evaluation research goals,
- A strong understanding of sampling methods and approaches and the ability to identify potential biases in a sampling approach and to develop control strategies for mitigating levels of bias, and a
- High level of past experience in conducting process evaluations of energy efficiency programs and in reporting the results of these studies.

Introduction and Key Issues

Definition

For the purposes of the Framework, a process evaluation is defined as: *a systematic assessment of an energy efficiency program for the purposes of (1) documenting program operations at the time of the examination, and (2) identifying and recommending improvements that can be made to the program to increase the program's efficiency or effectiveness for acquiring energy resources while maintaining high levels of participant satisfaction.*

This definition includes an assessment of the products and services provided by the program or the ways in which the program is designed, operated, and delivered. This definition also includes an assessment of the program's approaches to engaging and interacting with the target markets or market actors for the purpose of improving the program's cost-effectiveness. However, it should be noted that this definition focuses on identifying improvements or modifications to a program that directly or indirectly acquires energy impacts in the short-term (as in resource acquisition programs) or over the longer-term (as in market transformation programs) as a result of program actions. This definition specifically excludes the assessment of energy programs for purposes other than increasing the efficiency or effectiveness of the program to acquire energy resources, either directly or indirectly. For example, this definition excludes conducting management audits or evaluations for the purposes of supplementing a financial audit of a program unless these examinations, at least in part, are conducted for the purposes of reducing the net cost of acquiring the energy impacts.

Why These Two Goals are in the Definition

Essentially this definition has two components, documenting program operations and assessing programs for improvements that increase energy savings and improve cost-effectiveness. The primary purpose of the documentation component of this definition is to provide a baseline description of the program operations and processes and to compare the design and operational practices with the program theory. It is also important for the

consumers of process evaluation reports to have a full understanding of the current program processes and their importance in the program theory relationships prior to considering recommendations for change. Without the documentation component in the definition, it is possible for a process evaluation to report recommended changes to the program without a solid presentation of how the program operates, how operations support the program theory, or a description that allows for a full understanding of what it is that will be changed if the recommendation is carried out.

While most process evaluations focus their effort on the second of the definitional objectives, the first objective is an important part of the evaluation process. The first objective is important because the formal program plans and operational descriptions and manuals seldom provide an accurate or detailed reflection of how a program operates in practice. This is usually because of one or more of the following conditions.

- Program planning and presentation documents seldom describe program operations at a level of detail that allows others to understand the operational practices that make it successful or capable of improvement. These documents are typically prepared for the purpose of proposal presentation, for program delivery contracting, for providing documents for public consumption, or for sharing internally with program managers and staff for familiarization or training purposes. Seldom are these documents prepared for the purpose of sharing detailed operational procedures or systems with other stakeholders.
- Most programs evolve as they move through the implementation process. Even if program documents were originally developed with detailed step-by-step implementation descriptions, they are typically outdated in terms of their content and accuracy. This is not a criticism of energy programs or of their descriptive documents. Rather it is an acceptance of the understanding that programs evolve as they move through the design and delivery process and this evolution occurs on an ongoing basis.
- In some cases program documents will describe designs and operational procedures that are officially part of the program process, and may even be reflected in a detailed operational manual, but for one or more reasons these activities are not replicated in the implementation practice. These planned practices can be, and often are, inadvertently or purposefully dropped or modified along the way. In many cases program delivery problems develop that act to reduce a program's efficiencies or its effectiveness because program operations or implementation efforts do not match officially approved designs or procedures. This is different from the second issue stated above in that these are not purposeful evolutions in the program designs but are typically non-performance issues.

As a result of these three considerations the process evaluation serves as a valuable tool to describe the operational conditions, systems, and procedures at the time of the evaluation that have a direct impact on the ability of the program to accomplish its goals.

The process evaluation provides a method for historical preservation and presentation of the detailed operations of a program that are being assessed in a way that others can understand how the program works and how program modifications are expected to influence program results.

While the documentation process is important for programs funded with Public Goods Charge funds, there is a second important benefit associated with the documentation process that directly serves the process evaluation function. The process of documenting the program's operations and processes also provides the foundation for comparing how the program is operated in relation to the concepts presented in the program theory. The program theory presents the cause and effect rationale supporting the program's designs and operational practices.¹⁷⁵ As a result, the process evaluation should find that the program is designed and operated in a way that is consistent with the program theory. The documentation process provides the information needed to determine if the program's operational practices are consistent with the program theory or if there are significant disconnects between the program theory and the way the program is operated. When the process evaluation finds that a program has operational components that are inconsistent with the program theory, the process evaluation can assess if the program theory needs to be modified or if there are aspects of the program that do not appear to be an important or relevant component of the program. Some evaluators argue that this is the most important focus of the process evaluation, especially during the early implementation period when operational practices are being tested and refined.

Certainly, the primary reason for conducting process evaluations is to identify and recommend changes in a program's operational procedures or systems that can be expected to improve the program's efficiency or cost-effectiveness. These recommendations need to be developed so that they support the program or the program's operational practices consistent with the program theory or with a recommended change to the program theory. Indeed, this goal covers a lot of territory, and includes issues internal to the program's operations (such as management systems and procedures) as well as issues external to the program (such as issues associated with customer acceptance and levels of customer satisfaction). As suggested in the *2001 Framework Study*,¹⁷⁶ a process evaluation also examines the extent to which a program is meeting or can meet a customer or participant need. Programs that are not capable of meeting a customer need cannot be expected to be successful from the perspective of the participating customer, and as a result, impact goals may not be achieved.

The process evaluation takes on the challenge of evaluating most, if not all aspects associated with the design or operations of a program in order to improve the energy resources acquired (directly or indirectly) by that program. The process evaluation can examine a variety of issues. These often include:

¹⁷⁵ See Chapter 4: Evaluation Overview and Issues, for information on the use of program theory in the evaluation process.

¹⁷⁶ *A Framework for Planning and Assessing Publicly Funded Energy Efficiency*. (* Sebold et al. 2001). See the work written by Dr. Jane Peters.

- Program design, goal setting, common vision, and the improvement process,
- Program staffing, staff skills, training, management and operations,
- Program information and information support systems,
- Program targeting, marketing, and outreach efforts,
- Program goal attainment and implementation processes and results,
- Program theories, theory assumptions, and key theory relationships - especially their causal relationships,
- Program timing and timelines,
- Participant satisfaction (both overall and individual components that interact with the participant in order for these to be assessed),
- Quality control procedures and processes,
- Reasons for low participation rates,
- Reasons for overly high free riders, or too low a level of market effects, free drivers or spillover,
- Use of new practices or best practices, and
- Intended or unanticipated market effects, among others.

The Need for Process Evaluations

The need for process evaluations of energy programs began to be recognized in the late 1970s and early 1980s when the United States Department of Energy (USDOE) began developing energy conservation and energy efficiency programs to help stretch U.S. energy supplies. Several of these early programs were experiencing customer participation and acceptance problems. As a result, the USDOE and the state agencies implementing these programs began to conduct evaluations that focused on the reasons associated with low participation or customer acceptance problems. Examples of these problems included energy efficiency loan programs that were targeted at customers with poor credit or who do not believe in borrowing, or energy audit programs that provided audit reports that were not user-friendly and were, as a result, not read or acted upon by program participants. The USDOE and state agencies learned that they needed to research program designs and operational systems to make sure they were appropriate for the market in which they were targeting. In these early years, energy program process evaluations were typically incorporated into other evaluation activities such as impact evaluations or market research studies about a particular product or service. As the energy program field matured, program designers, managers, policy makers, and regulatory managers began to see the value in conducting process evaluations to improve program designs, to acquire more participants and increase the attainment of cost-effective energy savings. In the mid 1980s utility companies began to offer a wide range of demand side management (DSM) programs. These program providers quickly recognized the benefits of the process evaluation and began to focus evaluations specifically for the purpose of assessing the design, operations, performance, and customer interaction and acceptance characteristics of their DSM programs. In many cases these evaluations were in response to regulatory agencies needing to be assured that programs were well-designed and implemented.

By the early 1990s process evaluations had become a widely accepted and desired evaluation research tool within the energy program industry. In recognition of this fact the Electric Power Research Institute (EPRI) published the *DSM Process Evaluation Guidebook to Current Practice*.¹⁷⁷ This publication presented and described how the process evaluation fit within an integrated program evaluation system and demonstrated how process evaluations were instrumental in the program design and redesign process leading to higher quality, more effective programs. According to the EPRI Guidebook, the primary purposes of the process evaluation are to:

- Improve program performance with respect to internal administration, promotional practices, program delivery, incentive levels, and data management,
- Provide information to regulators and other interested parties that [energy programs] are being implemented effectively and modified or refined as necessary,
- Provide a means of improving customer satisfaction and identifying market threats and opportunities, and
- Provide a means of contributing to industry-wide knowledge in order that other providers may improve their programs.

These goals are consistent to the process evaluation goals presented in an earlier published document by Ben Bronfman and Jane Peters in 1991.¹⁷⁸ In this document the goals of the process evaluation are to:

- Improve program implementation efficiency,
- Assess market segments and targeting of specific segments,
- Improve the quality of the measures installed,
- Identify program design issues,
- Providing an accounting of program progress, and
- Examine special issues (measure life, program comprehensiveness, etc.).

These goals are just as important today as they were in 1992. In California, energy resource suppliers, policy makers and regulators recognize the need for reliable program evaluation information as a tool to be able to more fully understand the potential energy supply and supply capabilities available through the State's energy programs. The State of California has adopted a policy decision that energy efficiency and renewable energy resources will contribute to lowering future per capita demand increases in California. As a result, the process evaluation has become an important tool in helping California policy makers understand the ability of energy programs to provide reliable energy resources and understand which programs and program operational and design practices can most effectively meet this goal. In California, energy programs are essentially a part

¹⁷⁷ *DSM Process Evaluation: A Guidebook To Current Practice*. (* EPRI 1992).

¹⁷⁸ *Process Evaluations of DSM Programs*. (Bronfman and Peters 1991), p. 139.

of a statewide effort to make sure the State has reliable cost-effective methods of meeting the energy needs of its households and businesses.

Process Evaluations are In-Depth Studies

The process evaluation must be conducted at a sufficiently detailed level of investigation to be able to make clear and specific recommendations pertaining to what aspects of the program's management, structure, function, and operations need to be changed. The process evaluation should provide managers the information they need to make improvements to their programs, ranging from major improvements to the program's product or product lines, program structure, program service, program operations, and program marketing, to making small adjustments to products, services or operational methods.

The nature of the process evaluation is that it is an investigation into the detailed operations of the program, and often includes investigating the customer's levels of interaction with the program, the product, and/or the service provided. For example, investigations into program operations focus on the structure, function, and operational details associated with the program. Investigations into customer satisfaction focus on customer acceptance and use of a product within the three primary characteristics that influence satisfaction, including the customer-product relationship, the product-provider relationship and the customer-provider relationship. While program managers may track overall satisfaction scores for management tracking purposes, a process evaluation will investigate much deeper into the satisfaction ratings and investigate the characteristics of the program, product, and/or delivery methods that influence satisfaction and identify specific program, product, and/or delivery approach changes that can increase satisfaction and at the same time cost-effectively accomplish the energy savings goals.

Evaluation Timing

Process evaluations can be conducted at any time within the program design and implementation process. There are, however, some key considerations for the timing of these evaluations associated with the Framework.

1. Design Feedback. In many cases, process evaluators can help programs be more effective or have more efficient operational systems before they are placed in the market. When process evaluation experts work with program designers and managers during the early development and implementation period, the evaluation staff can help identify potential problems associated with early program designs or operational practices. In addition, the process evaluation professional can work with program designers and managers to make sure they set up data management and operational systems that can operate to provide necessary reporting and evaluation data, and therefore be successfully evaluated during the process evaluation efforts. Evaluation experts have become accustomed to hearing the phrase "*we did not design the system to collect or trace the information you need to conduct your evaluation*" when they are not involved in the program design and planning process. Program designers and managers that work with their

- evaluation experts are better able to track and report their program progress and their programs are more easily, more efficiently, and more economically evaluated. This means that incorporating process evaluation feedback into the early program planning and implementation phases can be a very good investment for the program. It is almost always more expensive and more time consuming to modify program designs to collect the information needed to support the program evaluation effort after program data collection and tracking systems have been designed and are in operation.
2. Early Implementation Feedback. Early program evaluation feedback to the program designers and managers is an important component of any evaluation. If the process evaluation expert is brought into the program early, the evaluation expert can work directly with the program designers and managers to identify and agree upon an early feedback system that allows managers to be aware of early evaluation findings, and take corrective actions where appropriate before the evaluation report is finalized. These early feedback systems provide the program designer and manager with a method of ongoing interaction with the evaluation professional to make sure evaluation results are communicated to program management to allow for program improvements as early as possible.
 3. Scheduling the Evaluation Efforts. Process evaluations should be conducted for all significant energy programs early in the implementation process so that improvements to the programs can be identified and initiated over a period of time in which the results from the evaluation can influence the program being evaluated. While a process evaluation conducted near the end of a program funding cycle may help future programs (including extensions of current programs), the evaluation should be conducted in time to help the current program more effectively or more efficiently achieve its goals. This means that the process evaluation may need to be conducted after the program's start-up issues have been identified and dealt with by program management, or at which point the program is considered to be in a steady-state mode of operation. Typically, a new program requires about four to six months to move from the early planning and organization phase into the early market entry phase and on to a more normal, steady state operational phase. Typically this means that an in-depth process evaluation can be initiated in about the sixth month following program rollout. However, there may be a need to schedule the process evaluation over multiple years if the evaluation budget does not allow for a full process evaluation during the first year or if the program is implemented in phases such that process evaluations focus on each phase of the rollout. Likewise, the process evaluation can be scheduled to support a planned program redesign process or a mid-course reassessment. A part of this effort may be to compare the program's design and operations to the program theory and logic models. This comparison can assess if a program design or operational change can cost-effectively improve goal attainment or to see if the program theory or the logic model needs adjusting to reflect program operations. In this case the process evaluation may be conducted early enough to inform the redesign process and then later to assess the success of the redesign effort.

Process evaluations can also focus on and address different researchable issues during different phases of the program implementation and evaluation cycle. It is not necessary to plan process evaluations for all aspects of the program to occur at the same time. The benefit of scheduling and conducting process evaluations over the course of a program also have financial benefits in that the evaluation can be budgeted over the program funding cycle, rather than budgeted all at once. For example, the first year evaluation effort may want to focus on:

- Program design and operational systems,
- Program tracking and information management systems,
- Program delivery organization and staffing,
- Skill levels needed to implement the program,
- The methods and procedures used to target the outreach efforts,
- The marketing materials used to promote the program,
- Program operational efforts and their relationship to the program theory and logic model,
- The outreach efforts and the structure and content of these efforts, and
- Early program satisfaction and customer service experiences.

These issues can all be examined during the first year of the program operations. However, some aspects of the process evaluation may be more successfully assessed during the second or third year of a program cycle. For example, evaluations conducted after the first year might focus on:

- Assessing the relationship between the current program services and the needs of the market or participant,
- The program implementation system and its influence on customer perception of the program,
- The influence of the program on customer behavior and actions, and
- Field practices and their effects on energy savings achieved.

Most programs do not need a process evaluation every year of their implementation cycle. As stated earlier, new programs may want to undergo a process evaluation in the first year and involve the program evaluation staff early in the design process. For this reason, programs may want their process evaluation team on board and engaged during the early development efforts and for conducting the formal process evaluation initiated within the first year. However, if there have been previous process evaluation efforts for the program, the planning and conducting of the process evaluation needs to be informed by the past evaluation history. If the program has already had a rigorous process evaluation of the program's key components and the program has not significantly changed since that time, there may not be a need for a process evaluation for that

program. The process evaluation decision path discussed later in this chapter provides guidance for when and under what conditions process evaluations of energy programs should be considered.

Process Evaluation Activities

Process evaluations involve a wide range of activities in order to investigate the researchable issues associated with energy programs. These activities include, but are not limited to the use of:

- a. Reviews and tests of records, materials, tools, etc.,
- b. Interviews and discussions with program management and staff, implementing contractors, subcontractors, and field staff,
- c. Interviews and discussions with policy makers, key stakeholders, and market actors,
- d. Interviews, discussions, surveys and/or focus groups with participants and non-participants,
- e. Collection and analysis of relevant data available from third-party sources (e.g. equipment vendors),
- f. Field observations, measurements, and examinations, and
- g. Other activities as needed to address researchable issues.

In the following paragraphs we elaborate briefly on each of these items.

Examinations and Tests of Records and Materials

Examination of program records and materials will give the evaluator a better understanding of how the program is implemented in practice and allows for testing of the content and accuracy of the program records. As part of this assessment, a sample of customer files can be examined to test for content and accuracy of the information collected and to assess the effects of incomplete, inaccurate, or uncollected information on the ability to effectively manage and operate the program. This assessment can include a review of both participant and non-participant data as well as an assessment of the ability of the records to support management and operational needs.

Program electronic tracking systems and reports are important tools that program managers generally use to keep track of program activities and make sure that program activities, both internal and external, are proceeding on a reasonable schedule in a way that will achieve the program goals. An examination of these systems and reports can help the evaluator assess the strengths and weaknesses of the program operations and can also lead to the development of recommendations on how the tracking system and system reports can be improved. Likewise, this evaluation can examine the design of the database and database systems to assess the ability of the database to be modified, expanded, or upgraded, and to assess the ability of the tracking system reports to meet the needs of the program operations.

The program database is often a very valuable source of information. The database can be used to track, manage and report program marketing contacts, numbers of customers at different stages in the program participation process, and customers who start to participate in the program but then drop out. Databases can also be examined to determine the time that elapses between each stage of program implementation for a customer in order to identify barriers or roadblocks. Finally, program databases are a key source of names to draw from for subsequent evaluation interviews, focus groups and surveys.

Reviews of Management and Program Operations

The process evaluation can conduct a review of program operations and management efforts. For example, this review can include discussions with program staff detailing their efforts, activities and responsibilities, the steps they take to keep the program on track, and discussions with program managers and supervisors to review operational responsibilities, activities, and monitoring efforts. The process evaluation can then chart the management and operational systems and compare those to the program theory and logic models to see if program management and operational efforts are consistent with the primary program goals. Such reviews can help the evaluator assess what should be considered to improve the management and operations of the program.

Examinations of Customer Handling and Service Delivery Efforts

Customer handling and service delivery are core components of most energy programs. How customers are handled and the way in which services are delivered can lead directly to implemented projects and significantly influence satisfaction. Often customer handling and service delivery efforts directly affect the ability of the program to obtain participants who take the actions needed to provide the energy saved. A review of these procedures can include discussions with project staff, observation of interaction with the customer or participant, and conducting interviews or surveys with a sample of participants.

Assessments of Operational Tools and Procedures

For many programs the ability to accomplish its intended goals is influenced by the tools and equipment that are required to operate the program, including the use of software and hardware. A process evaluation can assess the program tools to determine if the tools and the operational procedures relating to the use of the tools are consistent with the ability of the program to accomplish its goals while maintaining strong customer satisfaction ratings.

Interviews and Discussions with Program Management and Staff, Implementing Contractors, Subcontractors, and Field Staff

Program management and staff can be a valuable source of information, as they typically know the program better than anyone. Interviews with lead program planners and managers, their supervisors, and a sample of program staff, including both central staff and field staff, can help the evaluator assess the program design and operations in order to make recommendations for changes that will improve the ability of the program to

cost-effectively obtain energy savings. In some cases it may also be useful to interview the administrative staff such as the staff responsible for program databases and the processing of incentive payments in order to support the evaluation.

These interviews can examine all aspects of the program's operations. This may start with each party's understanding of the program's goals and objectives. It is important to know whether different program personnel or stakeholders have a common vision for the program or if there are significant differences that may impact the program's success. The interviews can focus on the processes and activities of the key members of the program team and how their efforts support the program's goals and objectives. The interviews can be used to obtain detailed information on the program processes as practiced compared to the planned processes. Other subjects important to cover across different personnel include communications within the program, communications with customers, and communications with stakeholders. In addition the interviews can obtain impressions of the program's strengths and weaknesses and perceptions of the program's successes and the quality of work that can be compared and contrasted with those perceptions from stakeholders and participants. These interviews provide an opportunity to gather recommendations for program improvements from the interviewed personnel.

Interviews and Discussions with Policy Makers, Key Stakeholders and Market Actors

In addition to program staff, many other individuals are involved in a program including policy makers (e.g. CPUC staff), utility managers, key stakeholders (e.g. trade associations and tenant groups), and a variety of other market actors such as product manufacturers, distributors, and installation contractors. It can often be useful to interview a sample of these key players in order to obtain their insights into what the program is doing well, and what can be improved.

Interviews, Discussions, Surveys and/or Focus Groups with Participants and Non-Participants

One purpose of virtually all process evaluations is to understand the experience of the customers participating in the program in order to design program improvements resulting from those experiences. Program participants have valuable perspectives on what aspects of the program work well and what aspects represent barriers to participation or satisfaction. Detailed feedback from participants is also important for determining whether the customer's perceptions of specific program attributes and delivery procedures conflict or mesh with program design and management perceptions. Beneficial detailed feedback can include levels of satisfaction with the program and their participation experience and satisfaction with various elements of the program including satisfaction with the product(s), the organization, scheduling, educational services, quality of work performed, attitude of site staff, responsiveness to questions/concerns, level of savings achieved, etc.¹⁷⁹

¹⁷⁹ Hall and Reed suggest that a process evaluation include three levels of satisfaction measurements, focusing on the relationships between the participant and the product or service received, the relationship between the participant and the organization implementing the program, and the relationship between the product or service provided and the organization providing the program. (Hall and Reed 1997a).

In addition, customers who are familiar with the program but elect not to participate (service rejecters) are also an important source of information. These customers can talk about barriers to participation and methods for overcoming those barriers. Depending on the number of program participants and the depth of the information to be collected, data from participating and non-participating customers can be collected with in-depth interviews, focus groups, and/or structured surveys. Generally, the smaller the number of participants and the more detailed the information collected, the more likely that in-depth interviews should be used. Questions that can be addressed using CATI systems typically focus on issues that can be addressed using simple questions leading to a “check the box” type of response. When more complicated or more detailed information is needed, a professionally conducted in-depth interview may be the best choice. Regardless of whether interviews, focus groups or surveys are used, care must be spent developing the questions that will be asked in order to provide objective responses at the level of depth needed to develop detailed process findings and recommendations.

There is a great deal of literature on developing and conducting high-quality surveys, interviews, and focus groups. Unfortunately, it is easy for inexperienced people to think that anyone can design a survey, interview guide, or focus group guide. However, there is a large field of survey design research that quickly points to many easily overlooked problems in instrument and question design that can decrease the validity and reliability of the information collected, lowering the usefulness of the information collected. Many of these issues become the basis for the different threats to validity discussed in Chapter 12, Uncertainty. Evaluators should be familiar with this literature and the state-of-the-art practice to be able to design meaningful and reliable evaluation studies. (Recommendations for using these tools and their associated methods can be found in most graduate level survey research design and survey development design textbooks.¹⁸⁰)

Collection and Analysis of Relevant Data Available from Third-Party Sources

Data to support a process evaluation may be available from third-party sources or from sources of secondary data. For example, data on equipment stocking or equipment sales may be available from manufacturers or local distributors, or information on how markets work or how customers make decisions may be available from dealers and other vendors. Likewise, census data or market information collected or maintained by other organizations can be helpful in conducting process evaluation assessments. For example, data from the federal ENERGY STAR[®] program can be helpful in understanding the markets for appliances that may be pushed by a California energy program. Other sources to support a process evaluation may include purchased market information from one of the for-profit companies that sell market or customer data. Customer, market or program information collected by other energy program evaluations may also help support process evaluation assessments.

¹⁸⁰ Examples include: Chapter 10 of *Applied Social Research* covers an introduction for Survey Research (Rubin 1983); *Focus Groups: A Practical guide for Applied Research*, 2nd edition. (Krueger 1994); and *Focus Groups as Qualitative Research*, 2nd edition. (Morgan 1997).

Field Observations, Measurements and Examinations

As noted above, sometimes it is useful for the process evaluator to go into the field with or without program staff to observe program operations, customer interaction practices, and customer education procedures and techniques used. Programs that have substantial on-site interaction with the participant involving measure installations, for example, may need to employ on-site examinations to assess the quality of the work provided. On-site observations can also be used in assessing the on-site approach and procedures associated with the program's interactions, relationships, and job performance.

Sometimes important data is not collected by the program or by third-party sources, requiring the evaluator to collect data in the field. An example might be that the evaluator needs to conduct a quick survey on the retail price of eligible equipment at a sample of retail stores in the program's target area, or may need to observe the installation of equipment installed by program staff to assess the installation process.

Other Activities as Needed to Address Researchable Issues

While the efforts discussed above can provide a wealth of evaluation information, sometimes other research activities or tools will be needed. For example, to review the adequacy of field efforts associated with the installation of insulation it may be appropriate to acquire infrared thermal imaging equipment to confirm the amount of installation installed or the adequacy of the installation. Likewise, it may be desirable for the evaluation team to hire and train interviewers from the target neighborhood who have pre-established relationships with members of the target group and can gain access to the target group. The evaluation team may find itself needing to be flexible and creative in identifying approaches for conducting the process evaluation that meets the data collection and analysis needs of the research effort. The process evaluation professional must select the right set of activities to match the researchable issues to be addressed, the available budget and other available resources to the evaluation.

Access to Program Materials, Resources, and Personnel

In conducting the process evaluation, it is typically necessary for the evaluation team to have unrestricted access to the program management and staff as well as a wide range of records, reports and materials associated with the program implementation. Program managers need to understand that the process evaluation will consume a small amount of their time and potentially the time of their staff for interviews and discussions and to help locate and acquire the materials that need to be reviewed during the evaluation. Evaluation professionals need to understand that the program management and staff need to incorporate these needs into their daily operational schedules and commitments. It is often necessary for the evaluation team to schedule events with program managers and staff in advance, so that program staff can manage their program responsibilities to incorporate the needs of the evaluation team. Cooperation and support by the evaluation team, the administrative team, and the implementation team are conditions that are a part of every program evaluation effort.

In most cases where information is needed from the program, the evaluation staff can prepare a formal information request identifying the data needed. These requests typically detail the exact data needed and the purpose for requesting the data.

Making Decisions and Selecting Methods: The Process Evaluation Roadmap

This section presents the evaluation decision roadmap, a tool to help determine if a process evaluation is needed for a specific program. These decision procedures are presented as steps in the process evaluation roadmap.

In using the process evaluation roadmap, decision makers can move through each of the twelve steps described in the roadmap. The decision steps in the roadmap can be used to guide the development of the evaluation plan and help it focus on the issues to be evaluated. For example, the first step in the roadmap is associated with a determination if any of the program's designs or operational and delivery characteristics are new to that program. If the program employs new designs or operational practices, the process evaluation can focus on the effectiveness or efficiency of these "new" practices and assess their influence on the program's operations and their consistency with the program theory. Once the roadmap decisions are made, the evaluation plan can be structured to focus on the aspects of the program that extend from the roadmap's step-by-step decision process. In some cases, the roadmap's decision process may identify research needs that are greater than what can be conducted in a single process evaluation or fit within the available process evaluation budget. When this occurs, the focus of the evaluation plan will need to be prioritized and perhaps funded over multiple years within the program funding cycle, or delayed until a following cycle if necessary. In other cases the roadmap may suggest that a process evaluation is not needed within the current funding cycle. As indicated earlier, the process evaluation roadmap is constructed to be used to plan the evaluation effort within a single program funding cycle.

The process evaluation roadmap is designed to help program administrators and others make decisions about when a process evaluation can be expected to provide benefits to the program. In moving through the roadmap decision process it is important to keep in mind that program conditions change over time and as program conditions change it may be necessary to re-examine the roadmap and update the process evaluation decisions. For example, if a program is considered to be cost-effective or moving in a cost-effective direction during the early implementation period, the roadmap suggests a process evaluation may not be needed. As a result, the program administrator may not plan or budget for a process evaluation. However, if that program becomes less cost-effective as it moves through the implementation cycle, this decision may need to be reassessed based on the most current program information. The roadmap is designed to be used in an annual reassessment process.

In moving through the process decision roadmap it may be appropriate for a decision associated with a specific step to be overridden or revised by an organization other than

those associated with the administration or implementation of the program. The CPUC staff or their representatives may identify a program condition that justifies the implementation of a process evaluation to deal with a specific issue. In these cases the CPUC staff will notify the program implementer or administrator as appropriate, if conditions relative to their program warrant a process evaluation. In these cases the program administrators should plan a process evaluation to address the specific issue(s) identified by the CPUC staff.

Steps in the Process Evaluation Roadmap

The following figure (Figure 8.1) provides the process evaluation roadmap to help guide readers through the decision-making steps about when process evaluations can provide valuable information for California energy program policy decisions and to identify potential program improvements.

The reader will notice that there are a series of decisions along the path. In some cases, the results of a decision lead directly to a suggestion to conduct a process evaluation. In other cases a decision leads to an intermediate decision. This intermediate decision is based on the response to a question concerning whether a process evaluation can be expected to increase energy or demand savings or improve the program's cost-effectiveness.

The following paragraphs present the steps in the process evaluation roadmap.

Step One

Upon entering the process evaluation roadmap, the first step is to determine whether any significant part of the program designs, operational systems, or delivery characteristics are new, are innovative, or represent a change to the program or the program goals since a process evaluation for this program has occurred.

If any key components of the program designs, operational systems or delivery characteristics are new or are considered to be innovative then a process evaluation can be conducted to examine those parts of the program that are considered new or innovative. This evaluation can focus on assessing the effects of those new or innovative characteristics on the operations and expected success of the program and compared to the program theory to determine if the theory and the implementation efforts are consistent. For these reasons the roadmap shows a direct line from the decision concerning the “newness” or “innovativeness” of the program characteristics to the recommendation to conduct a process evaluation. These evaluations help support the informational needs of California policy makers to be able to more fully understand the program characteristics being evaluated and make decisions about their applicability to future programs in California or to make decisions concerning whether those characteristics should be repeated, continued or identified as a best practice.

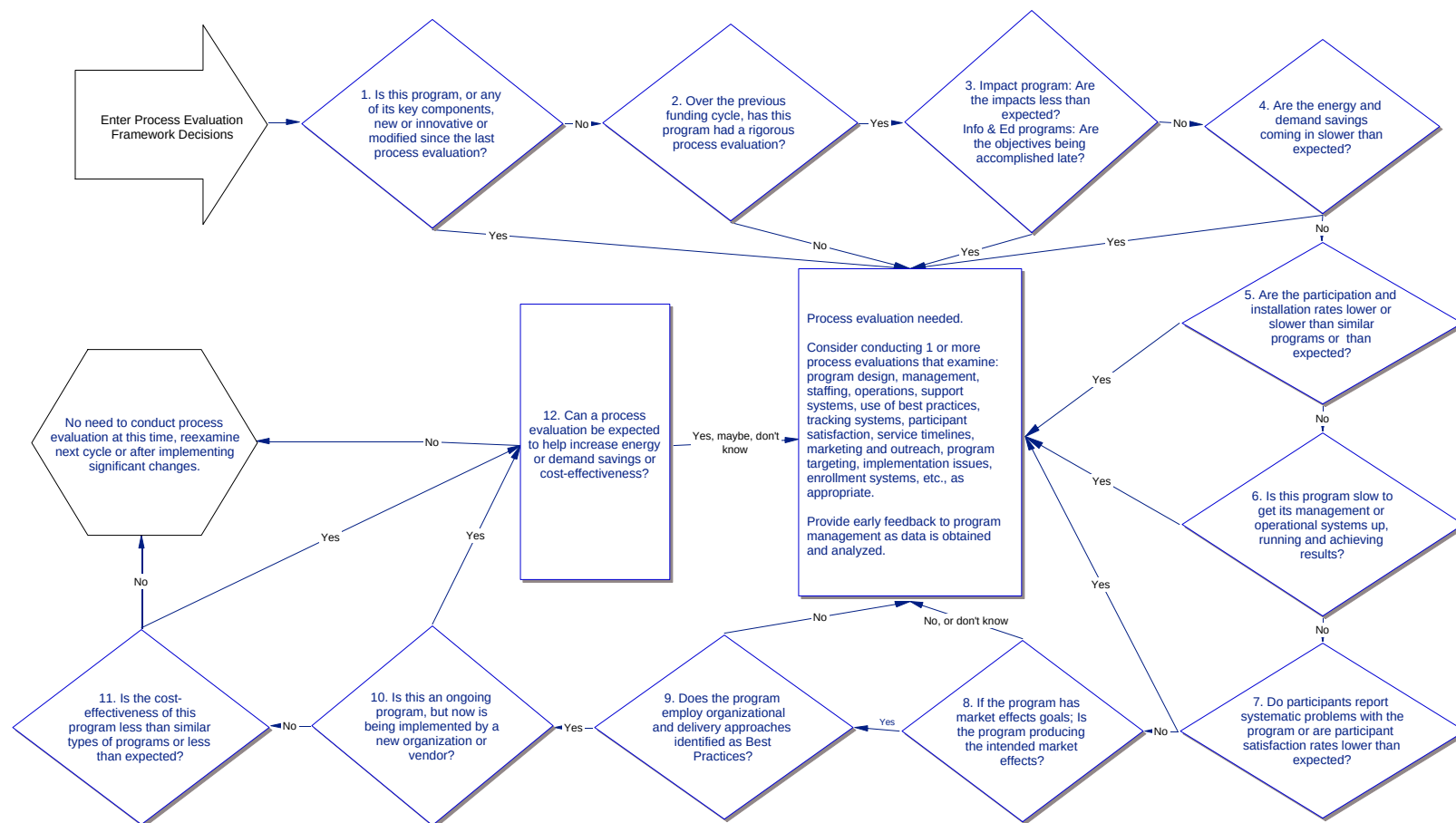


Figure 8.1: Process Evaluation Roadmap Decision Path

Step one also focuses on making a decision regarding whether the program has modified or changed any of its key design or delivery approaches pertaining to the management or operations of the program, including management systems, internal or external operations, or service delivery and customer interaction components. If the program has changed any of its design or implementation characteristics since the last process evaluation, then a process evaluation can be conducted to assess those parts of the program that have changed. This evaluation can focus on assessing the effects of those changes on the operations and expected success of the program and compared with the program theory.

If the response to step 1 is “no” there is nothing innovative about this program and “no” key component of the program has been modified since the last process evaluation, then the decision process moves on to step two.

Step Two

In Step two the roadmap looks at how long it has been since the program has had a rigorous process evaluation. If the program has not had a process evaluation of its designs, operational systems or delivery characteristics during the previous funding cycle, then a process evaluation is suggested for the current cycle. This evaluation would examine the program’s key design characteristics, its methods of operation and its program delivery systems, including its management, marketing, product mix and customer interaction processes.

If the response to step two is “yes” the program has had a rigorous process evaluation during the previous funding cycle then the decision process moves on to step 3.

Step Three

For programs with impact goals, this step seeks a decision to determine if the savings (energy or demand) are less than similar types of programs operating in California, per dollar of program costs, or if the energy savings are less than what is expected in comparison to the projected savings presented in the program implementation plans, the program theory, or in the program contract agreements. For programs that have had a previous net impact evaluation conducted on their program (or a very similar type of program) this decision can stem from a review of the net impact evaluation results compared to the program’s progress relative to the net impact goals. If the program does not have previous net impact evaluation results, then assessing the projected net impacts resulting from the program’s implementation progress can inform this decision. If the expected impacts from installed measures do not match the implementation plans of the program, then the program can implement a process evaluation to determine why impacts are lower or slower than expected. This can be especially important if the reason for low net effects is high free rider levels (for those that have estimated or measured free rider levels). In this case the process evaluation can look at what components of the program are causing the high free rider levels. In making the determination concerning whether the program is achieving less savings than expected, the decision maker should also review evaluation results from similar programs conducted in California and determine if their expected savings are lower, given their differing circumstances. We realize that this

will generally be a subjective assessment, especially in the early phases of a program, and decision makers must use their best judgment on using the information and how to interpret it, taking into account the conditions of the programs on which the comparisons are being made (i.e. technology differences, participant differences, weather differences, etc.). In some cases there may not be similar programs with which to draw a comparison, or there may not be impact information available on the programs being compared. This step asks for a best attempt to determine if the program is providing lower savings than what is or should be expected from the program at that point in the cycle. If the assessment indicates that the program is providing savings greater than or equal to that expected and/or to similar types of programs at similar points in their cycles, then the decision process moves to step four. If the assessment indicates that the program is providing savings that are less than expected or less than similar types of program (taking appropriate differences into consideration), then the program can undertake a process evaluation to address the reasons for the lower than expected savings and to develop recommendations for changes to the program that can be expected to improve the program's savings or cost-effectiveness.

For programs without an impact goal, such as information and education programs, this step asks if program objectives are being achieved consistent with the program implementation plans. For programs that have difficulty accomplishing their objectives, a process evaluation can be conducted to identify the reasons why objectives are not being met and what program modifications are needed to help bring the program back on schedule.

Step Four

While step three focused on lower than expected savings, this step assesses whether the energy savings (energy or demand) are being achieved at a slower rate than projected. If the project savings levels are coming in too slow and are not meeting expectations for the program, then the program should consider a process evaluation to assess the reasons for the slower than expected savings and identify recommendations that can be expected to speed the rate of achieved savings.

Step Five

This step is similar to and in many cases related to the decision associated with steps three and four. In this step the decision maker is asked to make an objective decision about the participation and measure installation rates for their program. This assessment should include a review of the program's projected accomplishments with regard to the number of participants served by the program and the number and type of measures installed through the program, and determine if these rates are different than what is needed to accomplish the stated goals in the program implementation plans or in the implementation contracts. If the participation and implementation rates are greater than or equal to what is needed to reach their projected accomplishments with regard to participants and installed measures, then a process evaluation to assess these conditions is not needed and the decision makers move on to step six. However, if the program participation rates or the measure installation rates are lower than projected, then a process evaluation can be conducted to identify the reasons for the slower than expected

participation or measure installation rates. This process evaluation should provide recommendations that can be expected to increase the participation or measure installation rates for the program.

Step Six

Step six involves a decision regarding the ability of the program to get its management or operational systems up, running and achieving results in a reasonable period of time. This again is a subjective assessment. However, programs that are slow to move from an approved program plan to an operational program, with management and operational systems in place and functioning in a way that allows the program to produce timely results, can consider conducting a process evaluation. This evaluation can focus on the reasons for the slow operations, and focus on providing recommendations for speeding up the development of the program's management and operational systems. Programs that have their management and operational system up and running and supporting the program in time for the program to function according to the program implementation plans do not need to conduct a process evaluation for this issue.

Step Seven

Step seven is placed in the roadmap to trigger a process evaluation if there are problems from the participant's perspective. These include problems with the program designs, operations, or services which lead to excessive participant complaints or reduced levels of participant satisfaction. While all energy programs can expect minor customer complaints from time to time, and can expect satisfaction scores that might be lower than desired from a very small fraction of participants, this step asks if there are systematic problems with the program from the participant's perspective that produce higher than desired complaints or lower than expected satisfaction levels. This again is a subjective assessment. However, if the program is having any significant level of participant interaction problems that stem from the design of the program, the methods of operation or the services received, a process evaluation assessing these problems and the reasons for these problems should be seriously considered. The evaluation should focus on providing program design changes, changes to operational systems, or changes to the customer interaction components that can be expected to reduce complaint levels and increase levels of satisfaction. Complaints or low satisfaction levels from one or two percent of customers may be considered reasonable; however, complaints or low satisfaction levels from five percent or more of the participants should be considered excessive. The goal of the CPUC is to provide energy resource producing programs that have high customer satisfaction rates, low complaint levels and cost-effective impacts. This step supports that goal.

Step Eight

Step eight is the final step in the roadmap that leads directly from a program associated decision directly to a recommendation to conduct a process evaluation. This step is for programs that have market effects or market transformation goals in their program implementation plans or in their implementation contracts. These programs can conduct market effects (transformation) evaluations via the Framework as indicated in Chapter 10 of this document. If this evaluation indicates that the market effects projected in the

implementation plans or the program is achieving program contracts, the program does not need to consider a process evaluation to address market effects issues. However, if the evaluation indicates that the program goals pertaining to market effects are not being achieved, a process evaluation that works with a market transformation or market effects evaluation can be conducted to address the reasons for the low effects and provide recommendations for how the program can be improved to increase the effects in the market, if possible.

Steps nine, ten, and eleven incorporate an additional decision in the sequence before a determination can be made regarding a recommendation to conduct a process evaluation. The added step in the decision process is a subjective determination regarding whether the program can be expected to increase energy or demand savings or increase the cost-effectiveness of the program as a result of conducting a process evaluation (see step twelve).

Step Nine

In Step nine, a decision is made regarding whether the program employs appropriate “best practices” identified in the 2004 National Best Practices Study pertaining to the design and operations of Public Good Charge programs, or in the services or service delivery approaches used by these programs.¹⁸¹ The CPUC desires that all energy efficiency programs use best practice approaches that improve program operations or service delivery whenever possible and appropriate (cost-effective). This decision rests on the completion of the 2004 National Best Practices Study (or other best practice studies) and the identification of approaches and designs that can be considered best practices. This study will provide a presentation of best practices as a resource to program designers, managers, administrators and contractors. This step in the roadmap decision process asks programs to assess these best practices as to what might be applicable to their program and whether incorporating additional best practices might help their program increase savings or make it more cost-effective.

If the program is not using the best practices identified for their type of program, then a process evaluation to examine the potential for use of the applicable best practice is suggested. This process evaluation can examine the current program designs, operations and customer interaction systems and make recommendations for the inclusion of the appropriate best practices. Again, this is a subjective decision process that may be overridden or modified by the CPUC if the CPUC determines that a process evaluation to examine the use of best practices is needed.

Step Ten

In step ten, the decision is based upon whether a program is being implemented by an administrator or contractor who has not implemented this type of program in the past five years within the state of California. If this is the case, then the process evaluation decision moves to the next level (step eleven) to decide if a process evaluation can be expected to increase savings or cost-effectiveness. Programs that are provided by administrators who may not be familiar with California’s special conditions (weather,

¹⁸¹ *National Best Practices Study*. (Quantum Consulting Inc. 2004).

markets, customer opinions and needs, consumer protection laws, past programs and program operations, etc.) may be at a disadvantage compared to others who are established and experienced with California's energy programs and markets. Organizations and vendors who have not provided the type of program they currently offer need to assess their operations and determine if they think their program can benefit from a process evaluation that examines their structure, operations, services and service delivery efforts to identify ways that the program can be changed to be more effective at achieving energy impacts in the California environment.

Step Eleven

This step involves a decision associated with the program's cost-effectiveness. If the program is operating at lower than expected levels of cost-effectiveness, then a process evaluation to examine the reasons for this performance is suggested. While most programs experience start-up and early rollout conditions that act to lower a program's cost-effectiveness ranking, these conditions should be overcome early in the program cycle. If program design, operation, or market conditions are not allowing the program to operate at their planned levels of energy impacts per dollar spent, then a process evaluation to address these issues should be considered. This decision involves a determination of whether the current program cost-effectiveness is lower than what is described in the program implementation plan or the implementation contract for the current period of the program. If during the course of the program's implementation efforts the cost-effectiveness falls below what should be expected from the program, then a decision should be made concerning whether a process evaluation can be expected to help increase energy or demand savings or help increase the program's cost-effectiveness.

Step Twelve

Step twelve is the final decision step in the process evaluation roadmap. In this step the administrator makes a determination (based on their expert knowledge, the projected net impacts of similar programs included in the CPUC's program portfolios, the portfolio of progress reports available to program administrators from the CPUC, and the program evaluations conducted in California in the past five years that are available on the CALMAC web site), if their program can benefit from a process evaluation. If the administrator does not think that a process evaluation will help increase their program's energy impact performance or their cost-effectiveness, then a process evaluation is not recommended. However, if the administrator determines that a process evaluation may help achieve greater energy impact or improve the cost-effectiveness of their programs, or if they are not sure or don't know if a process evaluation would help, then a process evaluation that focuses on assessing program designs, operational practices, technologies, and market and customer interaction practices is recommended to determine if the program can improve its operational practices or delivery efforts. Again, the decision process in this case is being made by the administrator of the program, but may be subject to revision by the CPUC.

Conducting the Process Evaluation

If any one of the above detailed decision steps identify the need for a process evaluation, it is recommended that the program budget and plan for the implementation of the needed evaluation by an independent program evaluator that meets the skills requirements presented in this chapter.

Process Evaluation Budgets

Typically process evaluations range in cost from a low of around \$10,000 to \$15,000 for small evaluations of limited scope, to an average cost of around \$30,000 to \$60,000 for evaluations focusing on multiple issues with some level of supportive field data collection and detailed assessment efforts. However, it is not unusual for process evaluations that focus on a wider range of issues involving customer surveys, interviews or focus groups with on-site examinations to run in excess of \$60,000. For these reasons process evaluations should be scheduled into the program evaluation planning and budgeting process along with other evaluation efforts. For programs that are funded over multiple year cycles, the process evaluation may need to be structured over the course of the program to minimize the financial impact of the evaluation effort in a given year. However, process evaluations should typically be conducted early enough in the program cycle to be able to use the evaluation results to improve the program.

In establishing the evaluation budget many programs will need to prioritize the process evaluation issues on which they should focus their process evaluation dollars. It is up to the program administrator (perhaps in consultation with their evaluation contractor) to prioritize the process evaluation issues to be addressed within their process evaluation budget. However, the CPUC requires that the evaluation plans and budgets be approved by the CPUC prior to their implementation. This process will include a review of the program evaluation plans, negotiations concerning the scope and focus of these plans, and budgets. The CPUC may instruct the administrator to change their evaluation plans to meet the evaluation goals, objectives, and priorities of the CPUC.

Chapter 9: Information and Education Program Evaluation

Preface

This chapter of the Framework focuses on energy programs that provide information or education services to their target markets and addresses how these programs can be evaluated. A broad definition of information and education services is used to define the types of programs and efforts for which this chapter is applicable, including advertising, public service announcements, education efforts, training activities, outreach efforts, demonstration projects, and other information or communication-based efforts. These programs may be targeted to either end use customers or to other market actors whose activities influence the energy-related choices of end use customers. Market actors that influence end use customer choices include installation and repair contractors, retailer staffs, architects, design engineers, equipment distributors, manufacturers, and others.

Typically, California information and education programs have one or more of the following general goals:

1. Educate energy consumers regarding ways to increase the energy efficiency of their buildings and activities in order to convince them to take actions that help them manage their consumption or adopt more energy efficient practices.
2. Inform energy consumers and/or other market actors about program participation opportunities in order to increase enrollment in these programs.
3. Inform energy consumers and/or other market actors about energy issues, behaviors or products in an effort to transform the normal operations of the market, so that the targeted market actors make more energy efficient operational, purchase or behavior-oriented decisions *without* direct program interventions or incentives.

This chapter of the Framework focuses on the evaluation needs of California's information and education programs and the evaluation decisions associated with these programs. The information presented in this chapter is especially important to information and education program administrators, and to evaluation managers and their staff who conduct these types of evaluations, but also to individuals responsible for the evaluation planning budgeting and oversight functions. This chapter is also suggested reading for energy regulatory and policy managers and portfolio managers. The chapter is also written to provide a level of subject familiarity to individuals responsible for cost-effectiveness assessments and other evaluation professionals who may need to support or inform an evaluation of information and/or education program effects.

Introduction and Key Issues

It can be argued that almost every energy efficiency program provides some level of educational and/or informational content. However, if the program has been created primarily as a conduit that leads participants into other programs or services, or it provides training and education on energy efficiency options to customers and other market actors, then the program should not be expected to meet the same cost-effectiveness requirements as programs that are offered expressly as a way of acquiring energy resources. That is, these programs may not be expected to immediately produce cost-effective energy resources if they are not designed with that intent in mind. Rather, information and education programs are typically designed to acquire energy or peak savings indirectly through changes in behavior after exposure to the information, over time (market transformation), or via increased enrollments in other resource acquisition programs. As a result, information and education programs are generally not expected to pursue the Framework's energy impact or market transformation evaluation paths. Instead, a unique information and education effects path is provided in the Framework for these programs. This path is an alternative to conducting impact or market effects evaluations, and is in addition to the appropriate process evaluations suggested for these programs.

As a result, information and education programs are expected to have an energy impact evaluation instead of an information and education effects evaluation only if they have energy acquisition goals on which their cost-effectiveness is based. Likewise, programs do not need an information and education effects evaluation if they have market effects goals and if their cost-effectiveness is based on achieved market effects. This treatment allows the information and education programs to have no more than two types of evaluations per program funding cycle depending on the specific objectives of the program.

The Framework structure is focused on individual programs. While following the information and education evaluation roadmap (decision path) will generally mean that only two types of evaluations will be conducted on each program, it does not preclude conducting additional evaluations on these programs or on groups of information or educational programs. There may be cases in which additional evaluation efforts are needed to understand the effects of these programs. For example, while individual information or education programs may not have program-specific market effects, these programs examined in clusters may have substantial market effects or energy impacts. It may be desirable to occasionally examine the energy or market effects of these programs as a group in order to better understand their combined effects on energy or on their target markets. In these cases, additional evaluations may need to be conducted as overarching studies. A discussion of these studies is presented in Chapter 15 of the Framework (Overarching Evaluation Studies).

Evaluation efforts for information and education programs need to focus their evaluation efforts on the activities that are appropriate for their specific type of program and program goals. For most information and education programs, there are two primary

evaluation paths that are suggested in the Framework. These paths are summarized in Figure 9.1 and are covered in detail within the relevant chapters of the Framework that deal specifically with each of the types of evaluation efforts suggested.

The following diagram indicates that information and education programs can be evaluated as appropriate for the specific characteristics of the program being considered.

The CPUC advises that programs should prioritize their evaluation efforts so that effects evaluations have the highest evaluation priority followed by process evaluations, and that energy impact evaluations and their associated M&V efforts are more important than market effects or information and education effects evaluations for programs with energy impact goals. This is a departure from the previous California evaluation practice for information and education programs and is presented here to bring attention to the need to consider both program effects and process evaluations for these programs. The suggested evaluation efforts are summarized below in Figure 9.1.

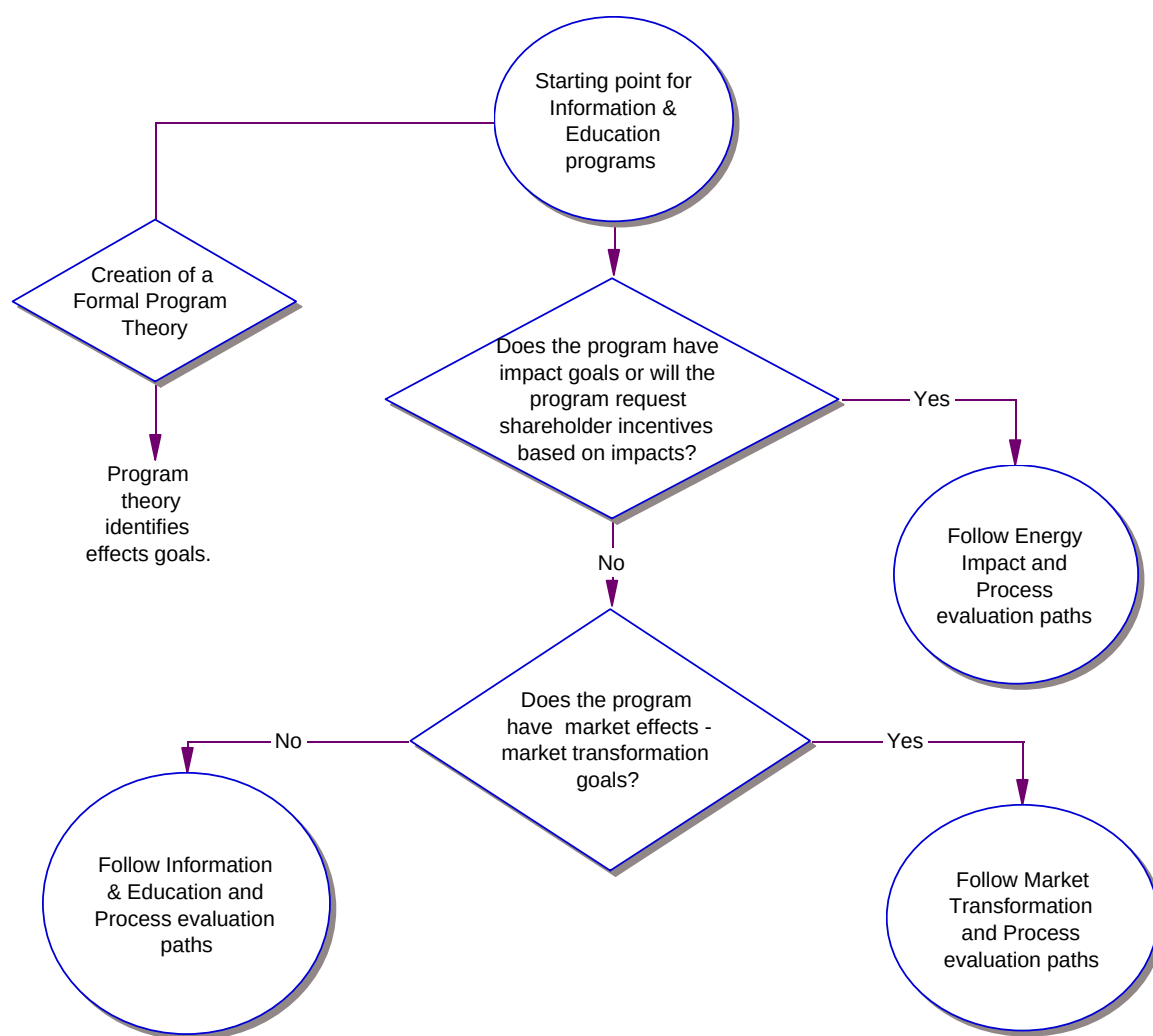


Figure 9.1: Overview of Information and Education Evaluation Efforts

The four possible evaluation paths for information/education program are described below.

- a. *Process Evaluation.* All information and education programs should consider following the process evaluation path presented in the Process Evaluation chapter of this Framework (Chapter 8). These evaluations help improve program design and implementation processes in order to improve their effectiveness or operational efficiencies. These evaluations also help document the operational processes that are employed so that they can be considered for use by other programs or as a best practice. All information and education programs should include a periodic process evaluation.
- b. *Market Transformation Evaluation.* Programs that have a goal of transforming a market or portion of a market, or of changing how a market operates, should follow the Market Transformation evaluation path presented in Chapter 10. These evaluations help document the amount of market change that has occurred as a result of the program's efforts and the effects of these changes on California's energy supplies. Programs with market transformation goals as their primary information and educational goals should consider conducting a market effects evaluation consistent with the market transformation evaluation roadmap.
- c. *Impact Evaluation.* California policy makers need to know how effective specific information and education programs are at increasing enrollments into the impact programs. Programs that are designed to refer energy consumers to other programs having resource acquisition or procurement goals need to coordinate their evaluation activities with the evaluations of the programs to which they are referring customers. It is important to assess the degree of success that information/education programs have in increasing customer awareness, enrollment and participation in the targeted resource acquisition programs. This type of evaluation typically takes the structure of including questions in the evaluation surveys and interviews of the resource acquisition program evaluations to determine the different ways participants hear about the program, which information sources helped guide the participant into the program enrollment process, and includes questions that allow the evaluator to determine the impact of the different information sources on the decision to enroll in the program. This activity usually involves adding researchable issues to the impact evaluation plans of the resource acquisition programs that deal with the success of the information or educational program's ability to influence customers to participate in the impact program.
- d. *Information/Educational Effects Evaluation.* If the program's goal is to lead customers or targeted market actors to take actions that reduce energy use (in the short-term or in the longer-term), and is classified as an information or education program (rather than energy impact, resource procurement, or market transformation), then the administrators should consider following the information and education effects program evaluation path described in the remaining sections of this chapter. The primary focus of this evaluation is to document the degree to which the program is achieving its desired effects within the markets targeted by the program.

Skills Required for Information/Educational Evaluations

Evaluations of information/education programs should be conducted by individuals or organizations that understand the relationships between providing information or training to different types of energy consumers and other market actors and the effects of that information on the behaviors of the individuals exposed to these services. Information and education program evaluators should be trained and experienced in conducting social science research specifically in the field of energy efficiency programs. Evaluators need a strong understanding of the cause and effect relationship between exposure to energy technology and behavior information and possible actions, and of the ways to test the causal processes involved.¹⁸² Evaluators should also have a strong understanding of the application of experimental and quasi-experimental research designs on human populations and be skilled at establishing evaluation approaches to test for net program effects.

Identifying the Issues

Information and education programs play a crucial role in the promotion of energy efficiency practices and measures in California. A significant number of programs fall under this classification, and information/education activities are significant parts of several resource acquisition programs. Consequently, documenting the effects of information and education programs and activities is an important component of the Framework. These evaluations help identify which information and education programs are meeting their goals and the degree to which those goals are met. It will also identify those programs that excel in the accomplishment of their goals. The primary purposes of the information and education program evaluation efforts are to:

- a. Provide information on the effects and effectiveness of the programs in motivating customers to either take efficiency actions on their own or to increase the possibility these customers will access or participate in other efficiency programs in order to determine which information and education programs are helping to meet California's current or future energy supply;
- b. Identify programs that need to be modified or improved to be more effective;
- c. Help identify best practices in the energy information and education program practice so that these practices can be documented, shared and replicated;
- d. Provide a system for documenting the accomplishments and the benefits received from spending Public Good Charge or energy procurement funds;
- e. Help policy makers and resource planners determine which program services to fund to help acquire future energy resources;

¹⁸² A study that provides an introduction to various methods of testing causal relationships and provides an analysis of methodological lessons learned after analyzing one of the country's oldest continuously operating energy efficiency programs is "How Can We Tell if Free Information is Really Transforming Our Market?" (* Conlon et al. 1999).

- f. Help resource supply planners identify a mix of energy resources that can be cost-effectively acquired to meet the energy needs of California's energy consumers.

Together the above reasons express the importance of and reasons for evaluating California's information and education programs.

Making Decisions and Selecting Methods

Once a program has been identified as an information or education program designed to cause people to take actions as a result of exposure to the program, the evaluation team needs to review the program plans, and particularly the detailed program theory associated with the program to identify the potential causal relationships between exposure to the program efforts and actions taken as a result. The program theory should describe in detail how the program promotes, changes or influences energy efficiency decisions. (See program theory discussion in Chapter 4: Evaluation Overview and Issues.) Typically this relationship is expressed through a formal program theory that may be linked to an implementation theory, a market operations theory, or to a program logic model.¹⁸³ Through the program theory process, program providers should describe how their program provides customers with information, education, and/or is being used to drive customers to other energy efficiency offerings or to take actions. Evaluators should then establish their evaluation approaches to test these relationships and effects to document and quantify actual effects and to confirm theoretical causal relationships. However, evaluators should not rely solely on the program's version of the program theory or the accompanying logic model to plan their evaluation efforts. Typically evaluators can identify other potential alternative causal relationships not reflected in the program theory or logic models for why a specific program effect may or may not be achieved. Evaluators should use the program theories and logic models as guides in the evaluation planning process rather than focus their evaluation efforts only on those relationships presented in the program theories and logic models.

Topics covered in the program theory should include:

- the educational or information subjects on which the program will focus and the efforts and activities to be undertaken;
- the specific education or information transfer methods and mechanisms that will be employed in the implementation process (audits, workshops, training classes, announcements, demonstrations, ads, etc.);
- the target market sector or sectors, including, as appropriate, market segments or sub-segments and the geographical market areas the program is designed to reach;

¹⁸³ One of the first uses of a logic model in the energy program field was in an evaluation of a children's educational project, "Detecting Behavioral Change from a Visit to a Children's Museum Energy Conservation Exhibit." This study found the program theory invaluable for structuring the evaluation approach. Evaluators were able to use the model for structuring the evaluation, making it possible to find the small changes made in energy consumption due to a children's museum exhibit. (* Peters et al. 2000).

- awareness, understanding, or knowledge of goals for target markets;
- the expected effects in terms of what the recipient is expected to do or accomplish as a result of the information or education efforts;
- the timeframes in which the expected results are to be accomplished;
- the barriers that the information or education must overcome to be successful; and
- the educational goals they are planning to meet within their program's market in terms of end effects.

A baseline study should also be conducted in order to understand initial energy efficiency issues and to establish the baseline from which program effects can be assessed. Issues to address in the baseline studies include:

- size and composition of target markets;
- pre-program awareness or knowledge levels;
- pre-program information and education sources;
- extent of exposure to and use of pre-program education or information sources;
- pre-program status of the target market relative to the intended results of the program; and
- relative pre-program adoption and behavior patterns.

The baseline study does not need to be program specific, but it should provide evaluation planners with enough detail to assess program-specific attributes by desired segments of the program population (i.e. income level strata, hard-to-reach population characteristics, geographic location, etc.). That is, the baseline should be well-understood so that the incremental change in the actions or behaviors of the target market can be estimated using accepted evaluation approaches for the specific target groups and sectors on which the program's efforts are targeted.

The timing of the evaluation effort is a critical consideration in structuring the information and educational program evaluation effort. To document the effects of the program in terms of actions taken by customers or other market actors exposed to the program's efforts, the evaluation will need to wait until the program has had the time to accomplish their goals to a level that can be measured. This may mean that the evaluation effort may need to be delayed until the program theory indicates that the targets of the program's efforts have had enough time to accomplish the intended results.¹⁸⁴ It is not unusual for an evaluation to be delayed for six months to a year or more. However, in some cases it is advisable to establish a time series evaluation approach in which the program's effects are examined periodically to identify the rate of

¹⁸⁴ This concept was applied in "Cinergy's Home Energy House Call (HEHC) Program: An Information Program That Changes People's Lives. The HEHC Program was evaluated from six months to two years after participation to determine how many households were implementing energy efficiency measures after an expanded audit of their home. (* Riggert et al. 1999).

change of the intended effect in the target population. A time series approach can be useful for documenting the rate of change, and if necessary help identify improvement opportunities that can be addressed in order to increase the rate of change.

Programs that inform market actors in a way that is expected to cause them to take specific actions will need to identify the target markets for their efforts in a way that permits evaluators to contact members of that market and assess the program's effects.¹⁸⁵ Broader consumer-based message efforts can in some cases be difficult or costly to evaluate due to the need to find those with recall. However, energy information messages broadcast over television or radio stations as part of a relatively large campaign can expect to be successful at finding and testing consumers for recall. These customers can be assessed in order to evaluate their reactions and behaviors to the message by using random-digit dialing and a sample-quota system for those with recall.¹⁸⁶

Smaller general advertising/information efforts, on the other hand, may find the cost to document the effects of the program to be prohibitive. However, policy makers should also understand that if the cost to identify and contact an individual exposed to the program's message is prohibitive, the message dilution factor (percent of a target market exposed to a message) associated with that program may be so low as to cause the program to have no significant measurable effects on the target population. For a broadcast message, an evaluator may need to contact hundreds of potential readers or listeners to find a few who were exposed to the message and who remember it to the extent that they can describe the influence of the message on their behaviors. It may not be practical to spend thousands of dollars in data collection costs to acquire program effects information from such a small group of individuals; however, it may be very practicable to contact potential viewers or readers in a specific market to identify the percent of the targeted population who recall the program's efforts. If recall is low, then policy makers need to determine if the recall rate is worth the program expenses.¹⁸⁷

To help identify the population reached by an information or education program designed to influence behavior, the program implementers should collect exposure contact

¹⁸⁵ An evaluation of a targeted information and education campaign is discussed in "Transforming Markets Through Education and Information: A Study of the Pacific Energy Center." This study examines the effects of the Pacific Energy Center's campaign to transform the commercial building market to make it more energy efficient. (* Reed et al. 1999). Full PG&E report: (Reed and Hall 1998).

¹⁸⁶ An evaluation of a large advertising effort was able to conduct this type of evaluation. This evaluation was particularly unusual in that the advertising effort was a concentrated three-month campaign to induce load-shifting behavior in the summer (among other energy saving tips). The 2002 evaluation included random pre and post surveys immediately surrounding the campaign. The surveys were designed to gather input to actually calculate peak load reduction. The results were used within a method to estimate peak load reduction, and large load-shifting impacts from the advertising effort were found. "Quantifying Load-Shifting Benefits from an Advertising Campaign." (* Engle et al. 2003).

¹⁸⁷ "Evaluating the Impacts of Education/Outreach Programs: Lessons on Impacts, Methods, and Optimal Education" provides a comprehensive overview of various types of advertising campaigns and the evaluation approach to identify the exposure and recall rates. (* Green and Skumatz 2000).

information on the people they reach whenever possible.¹⁸⁸ For example, if residential or commercial energy audits are provided, then contact information on the recipient of that service should be maintained and made available to the evaluation staff. Likewise if the program conducts door-to-door efforts within targeted neighborhoods or communities, the program should maintain detailed information on the specific program areas targeted so that individuals residing in the target areas can be identified via city directories, geo-coding capabilities, or field efforts that sample in the target area.¹⁸⁹ In many cases it will be possible to collect recipient contact data if the program is initially designed to do so. For example, if the program distributes energy efficiency information at a public event via a hand-out information packet, participants who receive the handouts can be offered an incentive to provide their contact information in order to participate in a free drawing for a desired incentive (for example, a drawing for \$100 each hour over a ten hour distribution period) in exchange for agreeing to be contacted by the evaluation staff. However, if an incentive to collect the information is used, designers need to minimize the influence of the incentive as a filter for the types of people who may provide their contact information. The ability to identify and contact individuals exposed to the program's efforts is often a critical component in the ability to evaluate an information or education program.

In some cases it may not be possible to collect contact information at the time the program services are provided. In other cases it may not be affordable to conduct the appropriate screening efforts within the evaluation process. In other cases, the cost of a formal evaluation may simply be too high for the size of the program being evaluated. In these cases, whenever possible, program evaluation efforts should be incorporated into the program delivery services themselves. One example of this approach is an evaluation questionnaire administered at the time of the exposure to the program for all or a sample of the market receiving the program service. In these cases, the evaluation effort should obtain program operations and delivery feedback information, but also, to the extent possible, obtain program effects or anticipated program effects information. In addition, there should be an attempt, as part of the real-time evaluation efforts, to collect participant contact information for follow-up evaluation contacts.

Primary consideration in determining whether an information and education evaluation needs to be conducted include:

- the ability to conduct the evaluation,
- the value of the information to feed long-term energy supply decisions,
- the cost of the evaluation for that specific program cycle, and

¹⁸⁸ An evaluation of a low-cost residential DSM program found that on-site visits were more likely to encourage the replacement of existing appliances with more energy efficient models. A complete contact database made this evaluation possible. See "Evaluating Educational Effects in Pacific Gas and Electric's Energy Savings Plan." (* O'Meara and Flanagan 1994).

¹⁸⁹ One study of an educational campaign done by Puget Sound Power & Light and an advertising firm utilized mass media and direct marketing in a campaign to push energy conservation. The evaluation shows unintended benefits and successful results, partially due to the known geographical area of the campaign. See "Conservation Advertising Campaigns and Advertising Effectiveness Research: The Right Combination to Solidify the Conservation Ethic." (* Auch and McDonald 1994).

- the anticipated value of the evaluation results toward longer-term program selection choices.

It may not be desirable to allow continued funding for program efforts that cannot document energy or demand savings nor information or educational effects through their evaluations.

Program Evaluation Design, Tools, and Examples

The purpose of this section is to aid in the selection and design of the evaluation strategy for information and education programs. Evaluators should examine the following list of evaluation topics for potential inclusion in their evaluation plan and identify evaluation strategies that allow for collecting and analyzing the data required for these topics.

Examples of information and education Program evaluation topics:

- Number and percent of customers reached, or made aware;
- Number and percent of customers reached that take recommended actions;
- Number and type of actions taken as a result of the program;
- Changes in awareness or knowledge by topic or subject area, by type of customer targeted;
- Customer perception of the value of the information and/or education received;
- Elapsed time between information exposure and action(s) taken by type of customer targeted;
- Attribution of cause for actions taken when multiple causes may be associated with the actions taken;
- Demographic, firmographic, or psychographic information as appropriate;
- Influence of program on dealers, contractors, and trade allies;
- Effects of program on manufacturers and distributors; and
- Identification of barriers experienced by program, and the development of recommendations for addressing those barriers.

The program administrator will want to incorporate as many of these research objectives in their evaluation plans as practical for the specific program(s) being evaluated and the available evaluation resources.

For each program, a detailed evaluation plan should be developed that incorporates the appropriate elements from the research topics listed above. The following types of data collection methods and tools may be appropriate for these evaluations:

- Surveys: telephone, mail, Internet, e-mail;
- Interviews: in-depth, small group, individual;

- Focus groups;
- Site visits, verification visits;
- Pre & post testing to gauge effectiveness of training initiatives;
- Follow-up efforts (e.g. six months after participation) to assess how training or information has influenced energy efficiency choices, actions or behaviors; and
- The use of comparison groups so that net effects of the program can be estimated.

Table 9.1 lists several types of information and/or education programs and an example of an evaluation issue that can be addressed when conducting an evaluation of these programs. An example of the various tools and strategies that can be used to address each of these issues are also presented in this table. This table presents a single evaluation issue for each type of program, not to suggest that the evaluation effort should focus on only one issue, but to illustrate an evaluation approach for addressing that issue. The evaluation approach that can be applied to a specific evaluation goal should be coordinated with an information and education program evaluation expert to design an evaluation approach appropriate for each program.¹⁹⁰

¹⁹⁰ An example of an education program evaluation specifically designed to work with evaluating “educational unit” programs can be found in “The Enduring Effects of an Elementary School Energy Education Program.” This evaluation of an elementary school program provided evidence of long-term and cumulative impacts. The program used participant/non-participant comparison surveys and other information to provide an estimated value for increased conservation behavior by students. (* Hanson and Siegal 1995).

Table 9.1: Information and Educational Evaluation Methods and Tools

<u>Type of Program</u>	<u>Issue to be Evaluated</u>	<u>Evaluation Tools/Methods</u>
Residential Energy Audit Program	Actions taken as a result of the program audit	Survey of program participants compared to a comparison group of non-participants.
Residential Energy Audit Program	Level of increased education or knowledge	Post-program knowledge survey battery compared to a matched comparison group of non-participants, or pre and post program questionnaires to measure differences compared to a comparison group.
School-based Student Energy Education Program	Level of increased education or knowledge	Teacher-administered tests/surveys of participants and non-participants, or pre and post testing of participants and non-participants.
School-based Student Energy Education Program	Actions taken as a result of the training	Teacher-administered tests/surveys of participants and non-participants.
ENERGY STAR [®] Promotional Program	Increase in ENERGY STAR [®] appliance sales due to advertising	Compare ENERGY STAR [®] appliance knowledge and sales to similar cities and/or stores in which the program is not provided.
Energy Efficiency Program/Services General Advertising Campaign	Increase in awareness levels and participation in energy efficiency programs and service offerings	Surveys of matched populations exposed and unexposed to the campaign or the same population before and after the campaign. Comparisons of information sources and program enrollment causes between areas exposed and not exposed to the campaign.

Training Programs

Programs that provide training to their participants should consider incorporating the following types of research objectives into their evaluation plans:

- Pre-program level of knowledge to compare with post-program levels,
- The specific knowledge gained through the program,
- Relevance and usefulness of the training as it relates to their specific needs and opportunities to use the information,
- Future opportunities and plans for incorporating the knowledge gained into actions or behaviors that provide energy impacts,

- Whether participants would recommend the training to a friend or colleague, and
- Participant recommendations for improving the program.

Tools used to evaluate training programs in general may include on-site pre/post testing of participants and non-participants, typically involving surveys or questionnaires that may be coordinated with participant satisfaction questions to understand the value of the training and how the knowledge gained during the training will be used. Follow-up post-program surveys of participants are also useful in order to determine knowledge retention issues and applicability or use of training.¹⁹¹

Programs with large training efforts or ones that are designed solely for training should have evaluation designs that are mindful of the rich literature and methods on evaluating training programs that are available from the larger evaluation community.¹⁹² An in-depth training evaluation might attempt to measure, report, and make recommendations on how the training meets its goals and objectives and how it fits into the overall goals of the program.

Kirkpatrick developed one of the best-known evaluation frameworks for classifying impacts of training programs. He suggested four levels of evaluation for training programs.¹⁹³ Kirkpatrick suggests that the impact/influence of the training can be measured, assessed, and examined for potential improvements across four levels of scope, including:

1. Reaction level - The satisfaction of the trainees (see additional information on assessing levels of satisfaction in Chapter 8: Process Evaluation);
2. Learning level - The level of knowledge gained (measured by comparing pre and post training test results);
3. Work Behavior level - How the knowledge has been transferred into workplace behavior changes; and
4. Organizational level - Where knowledge gained has been used by the trainee to influence organizational behavior changes.

These principles were successfully incorporated into an internal preliminary energy efficiency evaluation planning effort for the EPA's ENERGY STAR® Building benchmark training effort.

¹⁹¹ "The Enduring Effects of an Elementary School Energy Education Program" is an example of how evaluation tools and methods are used in an evaluation to determine the measurable effects of education on students' energy-related knowledge, attitudes and conservation practices. (* Hanson and Siegal 1995).

¹⁹² See: *Handbook of Training Evaluation and Measurement Methods*. (Phillips 1997); *How to Measure Training Results: A Practical Guide to Tracking the Six Key Indicators*. (Phillips et al. 2002); and *The Bottomline on ROI: Basics, Benefits, & Barriers to Measuring Training & Performance Improvement*. (Phillips 2002). Additional papers and information is available from the American Society for Training and Development.

¹⁹³ Originally published as four articles in the *Journal for the American Society for Training Directors*, then revised. (Kirkpatrick 1996).

There have also been several alternatives and revisions recommended to Kirkpatrick's initial four-level approach: a five-level approach by Kaufman; a four-level approach called CIRO by Warr, Bird, and Rackham, and a five-level approach developed by Phillips and Phillips.¹⁹⁴ The five-level approach by Phillips and Phillips incorporates evaluating the return on the investment of training dollars and may be particularly useful for energy efficiency training programs.

Steps in Deciding if an Information/Education Effects Evaluation is Required

The information and educational program evaluation diagram presented in Figure 9.2 below is a roadmap to help guide planners through the steps required to decide if an effects evaluation is needed for the current program cycle.

Step One

The first step is to determine if the program should be or is classified as an information or education program. In California this decision is made in the early program planning and design efforts and is usually coordinated with the CPUC. Typically an information or education determination is made if the program provides only information or instructional services and if the program does not, in itself, have energy impact or market transformation goals that must be obtained. The program is also usually classified as an information program if its primary goal is to refer participants to other energy programs that do have energy impact goals.

Step Two

The next step is to determine if an information or education program effects evaluation has already been conducted for the program as currently designed and implemented. If not, then a program effects evaluation is suggested using the information and education program evaluation processes described above.

Step Three

In general, if a program effects evaluation was conducted for this program as it is currently designed and implemented, but was conducted prior to the previous program cycle, then a program effects evaluation is suggested for the current program cycle. This requirement allows the tracking of the effects of the program at least every two program cycles so that the effects of the program can be tracked over time to determine if the program is performing as intended and is accomplishing the program goals.

Step Four

Step four includes a group of conditional decisions that allow for an effects evaluation to be conducted more often or less often, depending on past evaluation efforts, the comparison of the results from these efforts, and if the program has significantly changed

¹⁹⁴ All these approaches are discussed in (Phillips 1997).

over that period of time. There are a number of possible outcomes to this decision process. These are presented below in bulleted formats to allow the reader to see the different possible decision results. Following the diagram (Figure 9.2) will help the reader move through the decision process.

Potential results from following the roadmap are:

- Even if an information or education effects evaluation has been conducted during the past program cycle, and if multiple evaluations have been conducted in previous cycles, a new effects evaluation is suggested if the program's goals or methods have significantly changed since these evaluations were performed.
- Programs with multiple evaluations over the past two program cycles that have not had significant program changes in their methods or goals do not need to consider an effects evaluation if the results of these evaluations are comparable and the findings are within 5% of each other. However, if the results of the previous evaluations show differences greater than 5% across comparable metrics, then a new effects evaluation is recommended.
- If a program effects evaluation was conducted on this program in its current design (program has not changed), and this evaluation was conducted *within* the last two program cycles, then an effects evaluation for the current program cycle is not recommended.

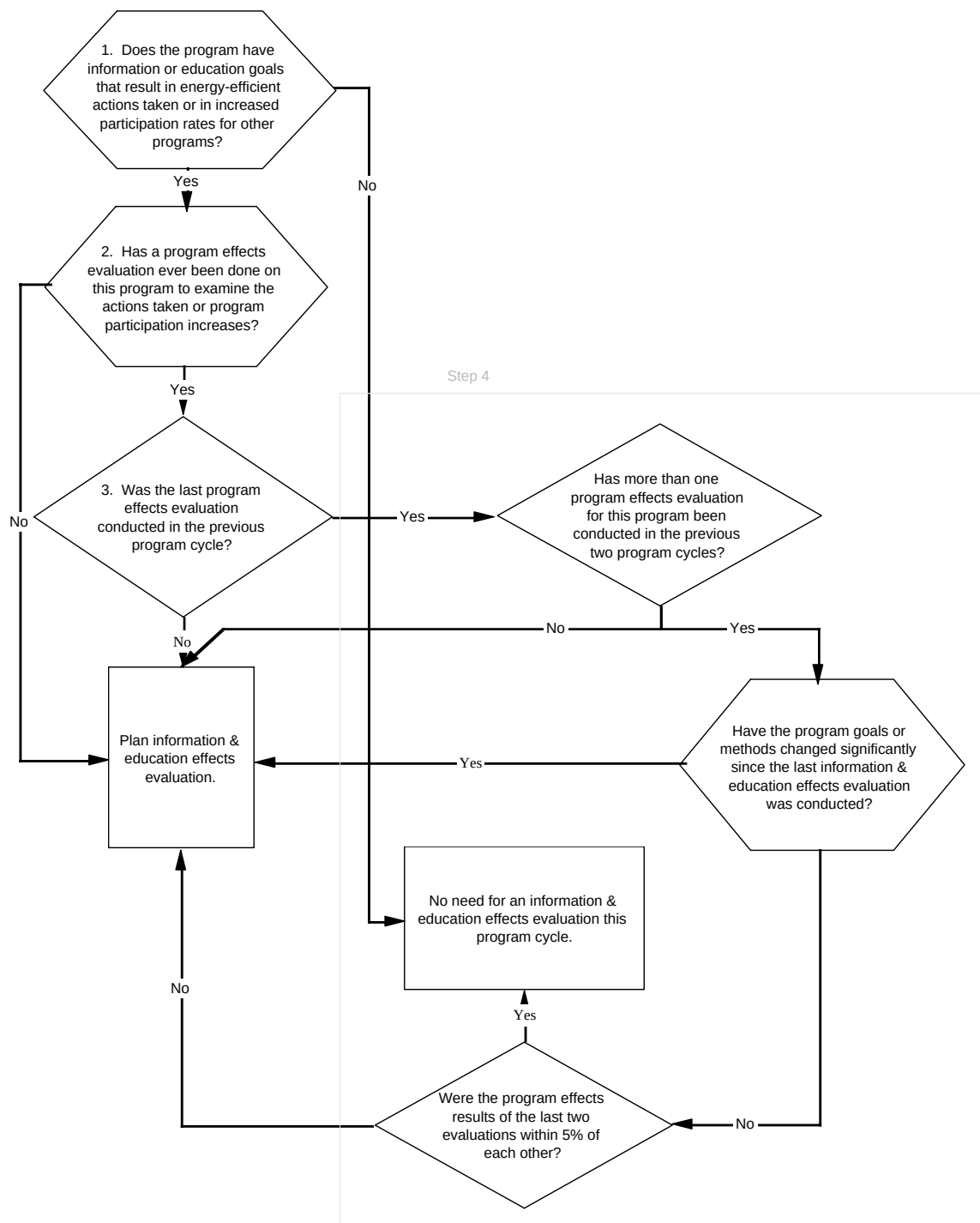


Figure 9.2: The Information and Education Evaluation Decision Roadmap

Chapter 10: Market Transformation Program Evaluation

Preface

This chapter of the Framework provides guidance on when a market transformation evaluation should be conducted and discusses the focus and activities associated with the evaluation effort. This chapter also discusses the general types of market-focused evaluations that can be conducted to support program planning and operational efforts, and at the same time help document market transformation (MT) program effects. This chapter also includes the market transformation evaluation roadmap. This roadmap (Figures 10.1 and 10.2) provides the step-by-step decision path to help program administrators and others determine when a market transformation evaluation should be considered. The information presented in this chapter can be useful for policy makers, as well as program administrators, and implementers needing an overview of the types of market-focused studies that are available for their consideration.

Four important principles are central to the recommended approach for conducting evaluations of market transformation programs. These are:

1. Market transformation program evaluations need to be conducted at the market, sub-market or niche market level rather than at the program level. This can mean conducting a market evaluation on a group of programs operating in the same market or conducting multiple market studies for a program operating in a number of markets.
2. There are a number of important conditions and activities needed to be able to evaluate market transformation programs, including an assessment of the program theory/logic model (PT/LM), a characterization of the market(s) in which the program(s) operate, the availability of baseline studies that provide a “starting point” for assessing market effects, market progress studies, and a causality assessment that examines the linkages between the program and the observed market changes (where the program-induced changes are defined as the market effects).
3. If a MT program evaluation is conducted to document program-created market change as part of an effort to estimate the energy impacts from a MT program, an energy impacts evaluation may still, in some cases, need to be conducted to verify the impacts achieved through the MT program.
4. A MT evaluation critically evaluates causality and sustainability. It is recommended that long-term market effects only be claimed under three conditions:
 - a. If the program theory specifically identifies the market effect(s) to be measured in the evaluation and provides theories supporting the causal relationship between the program’s efforts and the expected market effect(s),
 - b. If these efforts are supported in the program theory as being sustainable (i.e. last beyond the program period), or

- c. If sufficient evidence is provided through a MT evaluation indicating that the market effect(s) have a high probability of being the result of the program's efforts.

Skills Required for Market Effects Evaluation

This chapter provides the market transformation program evaluation roadmap. In doing so, it makes references to a variety of steps and types of studies for which it only provides a very brief description. It does not provide the in-depth discussions needed to learn how to perform these studies. Instead, the chapter focuses upon the basic outline of the studies needed and refers to other documents and articles as reference material. A working knowledge of this body of literature is recommended for policy makers, administrators, program designers, and evaluation managers involved in decisions regarding market transformation program's evaluation efforts.

The basic foundation for market transformation evaluation research is based upon research approaches developed within the larger field of program evaluation and market research. This body of research is found outside of the field of energy efficiency program evaluation. Most energy program evaluators will need to develop some familiarity with the adaptation of evaluation approaches in order to conduct market transformation evaluations of energy efficiency programs. Evaluators will also need to be aware of the current practice and history of the evolution of market transformation evaluation within the energy efficiency program literature. Training or experience with research dealing with microeconomics, diffusion of innovation¹⁹⁵ and adoption theory is critical to market transformation evaluation concepts and practices. At the same time, knowledge of program evaluation methods and issues (including the use of program theory and logic models and alternative hypothesis development and testing within the field of evaluation) are more important than academic economics for these types of evaluations. Depending on the nature of the market barriers or the program interventions, an understanding of social psychology and the associated social science research approaches can also be helpful. Market transformation evaluation researchers also need a solid understanding of how markets operate, what kinds of conditions and events move markets toward product adoption behaviors and what kinds of program-induced and non-program-induced conditions influence changes in market conditions that can be examined in a MT program evaluation.

¹⁹⁵ The seminal work in this area is *Diffusion of Innovations*, 4th Edition. (Rogers 1995).

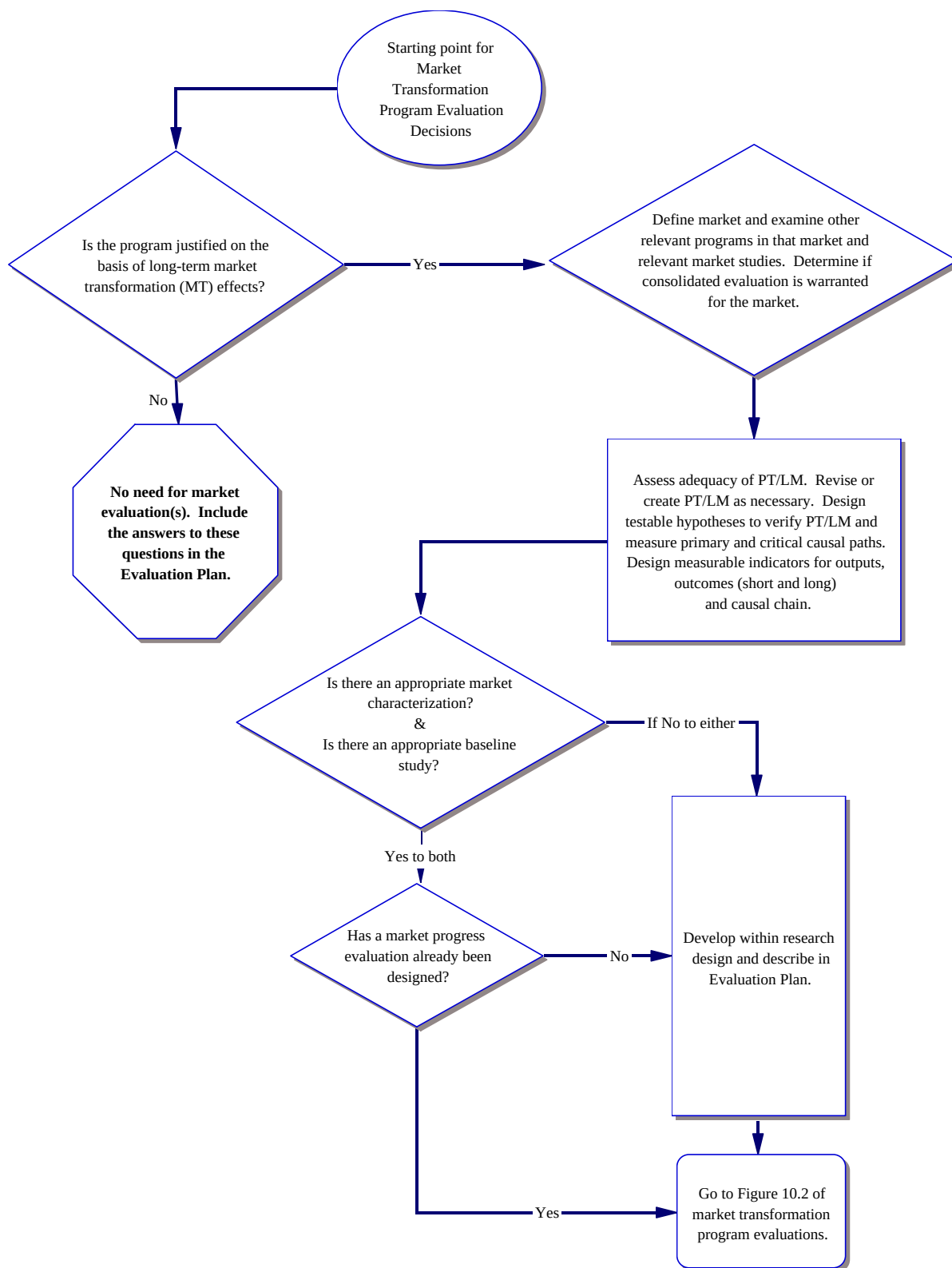


Figure 10.1: Market Transformation Program Evaluation Roadmap, First Half

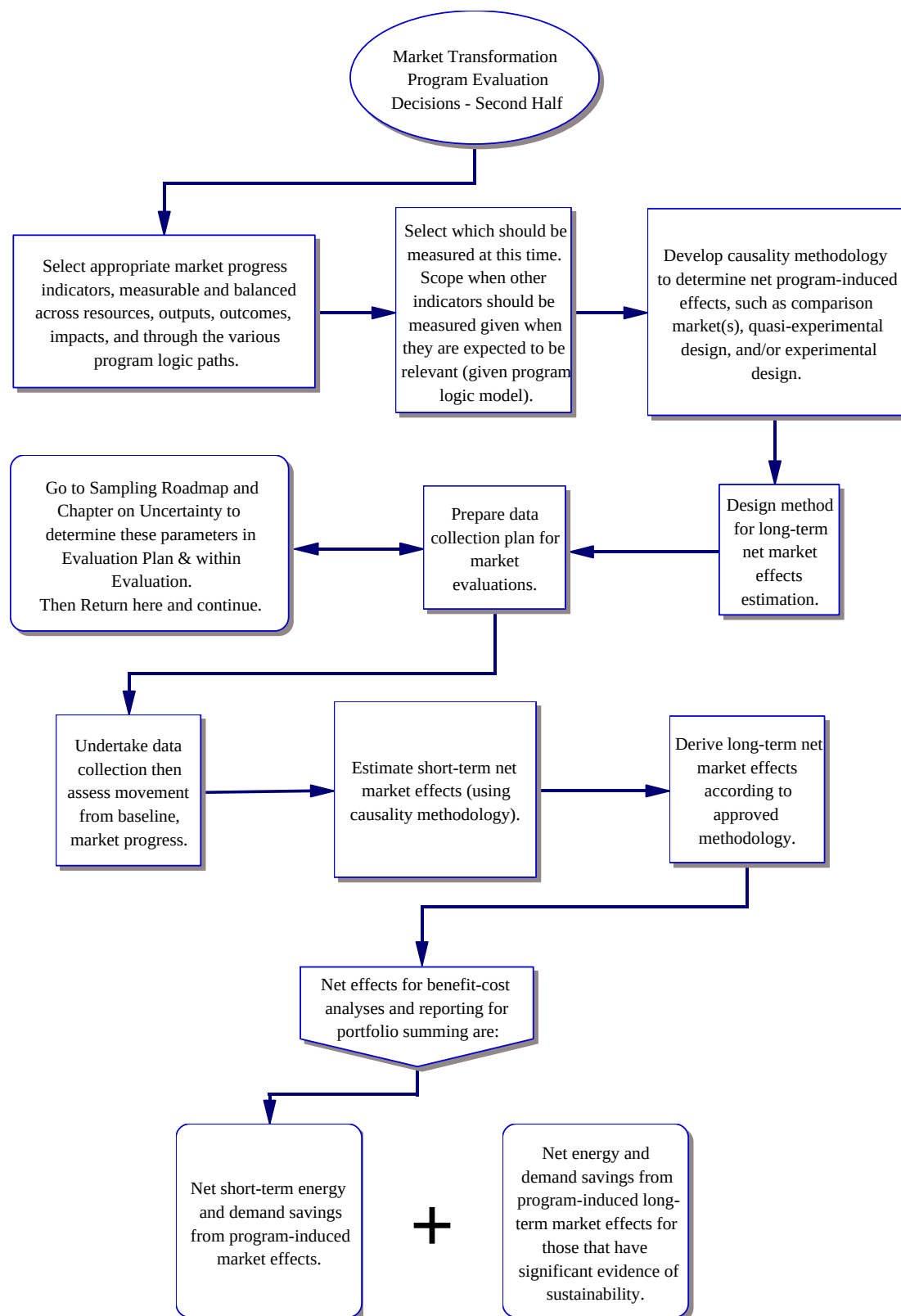


Figure 10.2: Market Transformation Program Evaluation Roadmap, Second Half

Introduction and Key Issues

As discussed in Chapter 5, Umbrella Roadmap, any program can claim to have an effect on the market in which it operates. Many programs have short-term market effects due to the interest they produce among trade allies, the learning that takes place during program influenced market transactions, or the contagious information exchange generated among participants and their peers. Short-term effects are best classified as “spillover” from the program intervention. As of 1998, few energy efficiency programs could claim long-term market effects, not because they are hard to measure, but because they are hard to accomplish. The main reason that (*most*) programs do not accomplish lasting market effects is because they are not *designed* specifically to address this goal (often because of regulatory policy directions given to program designers).¹⁹⁶

Without going into a full discussion of how to design market transformation (MT) programs to achieve long-term market effects, it is important to reiterate the elements that are likely to be found in a logic model for a MT program. These include a sustained effort, substantial non-energy benefits, allies in the market who see opportunities to capture benefits, and knowledge of the key communications relationships in the market.¹⁹⁷ Of these, a planned, sustained effort is a key indicator emphasized throughout Chapter 7 of the *2001 Framework Study*. There are multiple examples in the *2001 Framework Study* that illustrate how every program will have forward-moving “spillover” into future times, whether it is caused simply by education, rebates, or through a number of program efforts. The longer an effective intervention is placed in the market, the greater the penetration will likely be, and subsequently, the longer the carryover effect can be expected to influence the market. If long-term effects are forecast, then it is important to plan for long-term measurement and evaluation, regardless of the lifetime of the program.

Estimating the magnitude of the program’s effect on the market (and the ultimate outcome in terms of energy and demand savings) to include evidence of causality, and the sustainability of the market effect (the level of future savings after program discontinuation) is the toughest issue surrounding evaluation of MT programs. In order to determine the cost-effectiveness of a market transformation program, the evaluation needs to address the period over which the market effect will remain, the level of effect experienced in the market over time, the degree to which the program’s efforts caused the market effect, and the amount of energy savings provided by the effect.

¹⁹⁶ *Market Effects Summary Study, Volume 1* (of 3). (Peters et al. 1998). A summary can be seen in “Measuring Market Transformation: The 1997/1998 California Market Effects Studies.” (* Peters et al. 1998).

¹⁹⁷ “Wheat, Chaff and Conflicting Definitions in Market Transformation.” (Keating et al. 1998).

Market Identification and Evaluation Consolidation

As discussed earlier in this chapter, given the nature of their objectives, MT program evaluations need to be structured around markets rather than around a specific program in order to identify program-induced activities and the relevant market changes. Where there is a direct relationship between a program's activities and a market effect, the design of the evaluation may not be a difficult issue. However, it is not uncommon to find multiple programs targeting the same customers within a market or targeting different customers across a set of markets. Likewise, even a single program may target multiple markets.^{198, 199} In these cases, it may be appropriate to assess whether a consolidated evaluation covering multiple programs, or multiple markets, is desired.

An impediment to developing clear criteria for when to conduct consolidated evaluations is that there are no universally accepted definitions within the energy efficiency industry pertaining to what constitutes a program's market. For example, most observers would agree that a homeowner buying a CFL at a discount store and a manager of a manufacturing plant negotiating the purchase of new production equipment are participating in different markets. However, fewer program administrators or evaluators agree on whether a homeowner buying a CFL at a discount store, a small business owner buying CFLs by mail order, or a different homeowner buying a room air conditioner off the shelf at the same discount store, are participating in the same market.

Clearly, the question of whether two different programs are targeting the same market will often be one of degree, rather than one that lends itself to a simple yes or no answer. However, while there is no single universally accepted taxonomy of energy efficiency markets, there is more consensus as to the characteristics that tend to delineate and define markets. For example, most energy efficiency program professionals agree that the more the two energy efficiency related transactions share the following list of characteristics, the more likely it is that they are taking place in the same market.

- The same or similar product functions or categories of functions.
- Overlapping distribution chains.
- Nearby geographic locations.
- The same or competing manufacturers.
- Demographically or firmographically similar buyers.

¹⁹⁸ This is particularly likely to be an issue in a state in which market transformation is a secondary rather than a primary policy objective – a category that would seem at this time to include California. In a state in which market transformation is the primary policy objective, it is likely that programs will have been structured to correspond as closely as possible to individual markets, and thus there will tend to be a one-to-one correspondence between markets and programs. In a state in which market transformation is a secondary objective, programs are likely to have been structured based on other criteria.

¹⁹⁹ The converse issue, not discussed here due to space limitations, occurs when a single program targets multiple markets.

The decision to consolidate different programs for a MT program evaluation is often a subjective decision based on the relative extent to which the programs share market characteristics (such as those listed above). Likewise, the decision to address multiple markets in a single program evaluation should be based on the extent to which the transactions covered by the program share such characteristics.

Institutionally, it seems unlikely that any one evaluator or entity will have enough familiarity with the full range of programs being offered across the state to reliably assess whether there are other programs sharing enough of the market characteristics listed above that they should be consolidated for evaluation purposes. Thus it would seem that subjective decisions about consolidation would best involve a decision process considering multiple and perhaps conflicting issues. To address these issues it may be best to establish a process in which multiple issues can be addressed, considered and decided in a reasonable and timely manner. Such a process should be designed to operate within a more comprehensive evaluation planning process.

Assessing and Using Program Theory/Logic Models for Market Evaluation

This Framework emphasizes the importance of identifying, understanding and assessing the program's program theory and the associated program logic model (PT/LM). The Market Transformation Program Evaluation roadmap emphasizes the importance of understanding the goals of the program early in the evaluation planning process.²⁰⁰ If the program theory and the program goals explicitly identify market changes as a program effect or if it is clear that the cost-effectiveness of the program is predicated on counting long-term market effects, then the logic model should explain the key linkages between program activity and predicted effect. These explanations should include a clear description of the targeted market(s) and a detailed description of their operations, the market hypothesis on which the program activity and the expected effects rely, a description of the baseline condition that is expected to occur without program intervention, and the causal linkages that lead from program activities to the accomplishment of the program's goals. A benefit of this type of description is that it forces the program planners to be explicit about what they are doing, why they are doing it, and what the results will be. Without this information it is difficult to understand how the program expects to accomplish its goals. For example, a general goal of "taking X actions to produce Y effect" provides little information that explains the path between program actions and desired results.

Because of the importance of the PT/LM for MT evaluation, a critical first step in the MT program evaluation roadmap is assessing the PT/LM and revising it. If a PT/LM does not exist, it may be necessary for the evaluator to create one that is supported by the program administrator. The ability of the evaluator to design an effective evaluation rests on the

²⁰⁰ An introduction and references to the program theory/logic modeling literature can be found in Chapter 4 of this Framework. Examples of logic models are also provided in the Appendix to that chapter.

accuracy and comprehensiveness of the PT/LM. A poorly developed PT/LM can lead to a poorly designed evaluation plan unless the evaluator fully understands the program goals and activities and the operations of the market without the benefit of the PT/LM to guide the evaluation planning process. This means that the evaluator should not just grab an available PT/LM for the program or a similar program and use it to plan the evaluation. A critical assessment of the PT/LM or the creation of a PT/LM needs to be taken seriously by both the program team and the evaluation team. In this effort it is important to assess the PT/LM in terms of its thoroughness and accuracy in describing how a particular market operates and the various causal mechanisms lying within its patterns of operation. The PT/LM should be informed by a market operations theory.²⁰¹ The market operations theory is typically developed from a market operations analysis or a market baseline study. The market operations theory explicitly details the operations and operational linkages in the market that the program will need to influence.

With a PT/LM that accurately describes both the program logic and the operations of the market, an evaluator can move quickly to design a market evaluation.²⁰² The fundamental concept of evaluation is that it tests the program planning assumptions in the field.²⁰³ (See the *2001 Framework Study*, Chapters 3 and 6.) The logic model provides the assumptions behind the intervention and the program design. However, this roadmap suggests that if the goal of the program includes market effects, and if the market characterization, the baseline condition, and the market hypotheses and causal paths are not clearly presented in the PT/LM, then the evaluation should first clarify these issues with the program designers/implementers.²⁰⁴ Seldom can an evaluation provide enough evidence that the program induced the market effects without the causal relationships being clearly defined and the program testing these relationships in conjunction with the documentation of market changes that occur as a result of the changes caused by the program.

The logic model should provide a set of testable assumptions, some of which will provide market progress indicators and some of which may stretch the evaluation efforts out over a number of years.²⁰⁵ One of the roles of the evaluators will be to sift through the hypotheses that can be tested to identify those that are most crucial for assessing whether the logic model is accurate, showing progress of the program toward changing the market, measuring or estimate energy savings, and testing the long-term effects. The

²⁰¹ “Merging Program Theory and Market Theory in the Evaluation Planning Process.” (* Hall and Reed 2001).

²⁰² The PT/LM and market models will inevitably be refined as more is learned about the market through market transformation evaluations and program implementation experience.

²⁰³ In a pure acquisition program, the evaluation tests the assumptions behind the planned savings – the hours of use, the change in watts, the loading factors, all the costs, and the participation rate, as well as the free ridership. The assumptions being tested for a market transformation initiative more likely lie in whether the program intervention creates the expected outputs and short-term outcomes, and whether these lead to the desired long-term outcomes.

²⁰⁴ We note, however, that this does not relieve program designers of their obligation to attempt to ensure that the market hypotheses and causal paths in the program theory are clear to begin with.

²⁰⁵ The evaluation plan (as described in Chapter 5) should take this into account and indicate where a particular evaluation plan fits into the long-term evaluation plan for this particular market.

evaluation needs to test the assumptions upon which the logic model is built and whether the critical linkages actually occur (output to outcome to longer-term outcome to ultimate outcomes).²⁰⁶ Not all assumptions are equally important, and it is possible for an evaluation to get bogged down measuring activities instead of outputs or outcomes of a program. (See Chapter 5 in the *2001 Framework Study*.) Both Weiss and Rogers have recommended that evaluators consider testing the underlying mechanisms on which the program rests, which may include analyzing the psychosocial assumptions that can answer the question of why as well as how.²⁰⁷ Prioritizing the assumptions to be tested is important in order to control evaluation costs and focus evaluation efforts on those activities that provide the best assessments results.

There is extensive literature on the use of program theory and logic models in the program evaluation field outside of the smaller field of energy efficiency program evaluation. The authors recommend that this literature be used within the energy program evaluation field to support and evolve the use of program theory in the evaluation planning process. It is important that the evaluation team understand both the strengths and the weaknesses of using program theories as a data source for developing the evaluation design. Program theory's relevance to the evaluation of market transformation can clearly be seen by a classic evaluation textbook's treatment of the subject.

“Early data on measures relevant to the program’s long-term success are especially useful when the program has an extended time line. ... Theory-based evaluation calls for the collection and reporting of data on early phases that are conceptually linked to the program’s ability to produce long-term success. That is, the data are not just desired interim outcomes; they are the interim outcomes that are considered to be paths to long-term effects. They are part of the assumed casual chain.” ^{208, 209}

Once the desired testable hypotheses are isolated, the evaluator needs to deal with three key elements of the evaluation design. These are baseline, measurability, and causal arguments.²¹⁰ Each of these is further discussed later in this chapter.

There have been market transformation programs that intervene in the market by influencing key market actors to make them proponents for energy efficiency. Often, in these cases, the program is trying to create “agents of change.” There is significant theory from the diffusion of innovations literature that can be referenced when examining

²⁰⁶ Chapter 4 provides more information and references concerning program theory and logic models.

²⁰⁷ “Which Links in Which Theories Shall We Evaluate?” (Weiss 2000); *Program Theory in Evaluation: Challenges and Opportunities*. (* Rogers et al. 2000b).

²⁰⁸ *Evaluation: Methods for Studying Programs and Policies*, 2nd edition. (* Weiss 1998).

²⁰⁹ See also the Chapter on program theory in the largest-selling evaluation textbook: *Program Evaluation: A Systematic Approach*, 6th Edition. (Rossi et al. 1999); *Program Theory Evaluation: Practice, Promise, and Problems*. (* Rogers et al. 2000c); *Program Theory in Evaluation: Challenges and Opportunities*. (* Rogers et al. 2000b).

²¹⁰ There are plenty of other issues, such as timing, cost, persistence, and relationships among the various programs and their evaluations, but these are central to the design.

the program theory in this situation.²¹¹ Prior research across many different fields has identified the key characteristics of effective change agents and the steps they use to help create the desired change. This prior work can be used to establish evaluation criteria for assessing the MT program's selection of certain entities or individuals as agents of change. These criteria can also be used to evaluate whether the changes for which the selected agents are supposed to be the causal drivers are actually occurring within the marketplace and whether the changes caused by the agents appear to be due to the program. These types of evaluation efforts can serve as an example of how theory-based evaluation can enhance MT program evaluations.²¹²

Market Characterization and Baseline Measurement

Conceptually, market characterization and baseline measurement are two distinct activities, each of which plays a unique role in market evaluation. In practice, however, the boundaries between these two activities tend to be somewhat porous. For this reason market characterization and baseline measurement is discussed jointly in this section while exploring how the two activities are related.

Market Characterization can generally be defined as a qualitative assessment of the structure and functioning of a market, the primary purpose of which is to understand how the market operates in order to be able to effectively change the way in which the market functions.

Baseline Measurement can be defined as the quantification of key market indicators that have been or can be influenced by a program intervention. The primary purpose of the baseline measurement is to provide a basis for later comparisons of the status of the market after program intervention, in order to help assess the impact of the program.

Logically, while both market characterization and baseline measurement rely on the same kinds of market research data, there are at least three important distinctions between the two activities:

- Market Characterization efforts should generally precede baseline measurement. It is better to develop a qualitative understanding of a market before attempting to rigorously measure various elements and indicators of that market.
- From an evaluation perspective, the key purposes of a market characterization are to provide a starting point for assessing (or developing) the program theory, to help shape relevant baseline measurements, and later to help plan the market evaluation activities. The key function of the baseline measurement, on the other hand, is to help support an assessment of the impact of the program on the market.
- Baseline measurement, due to its critical role in evaluating market effects, should generally be performed by evaluators. Market characterization, in

²¹¹ See Chapter 9 in *Diffusion of Innovations*, 4th Edition. (Rogers 1995).

²¹² "Local Government Associations as Agents of Change." (* Megdal et al. 2000).

contrast, may be performed either by program evaluators, market research professionals, or by program staff.

In practice, however, market characterization and baseline measurement can often overlap. Budget constraints or small population sizes may force limitations on the size of baseline samples (such as the need to do interviews with the same market actors over a short period of time), thereby blurring the distinctions between the two activities. Likewise, time constraints may force evaluators to use the same data collection efforts for purposes of both market characterization and baseline measurement, thereby further blurring the boundaries. Nevertheless, it is still important to keep in mind both the differences and the similarities between these two MT program evaluation activities.

A complete taxonomy of all the elements that might be needed in a market characterization is beyond the scope of this project. What is known about a market should first be explored via secondary research. Next, the design of the market characterization study should then be based on the results of the usefulness of the findings from the secondary research. The research should be constructed so that together, the secondary and primary research should ensure an understanding of the current market structure and operation. Previous work in California provides some guidance on this issue. For example, a white paper prepared in 1998 under the direction of the California Board for Energy Efficiency listed the following types of information that should be considered in designing a market characterization or baseline study:²¹³

- a. A summary of the specific technologies, services or products being exchanged.
- b. A summary of the major market participants and the nature of the transactions and other interactions between them – including buyers, sellers and intermediaries.
- c. A description of how information concerning products and services flows through the market actors and the identification of key information hubs.
- d. A description of the distribution chain – i.e., the variety of paths that a product follows on its way from a manufacturer to an end user. A helpful tool here might be a product flow diagram.
- e. A description of the geographic boundaries of the market.
- f. A description of the circumstances and settings under which transactions tend to occur, including the sales practices and the market events that tend to result in transactions within the market (for example, a decision to remodel precipitating the purchase of a new C&I lighting system).
- g. Approximate estimates of the number of buyers, sellers and intermediaries in the market, as well as an order of magnitude estimate of the total annual sales of relevant measures and services.

Types of Market Indicators to be Studied

²¹³ *Proposed Guidelines for Conducting Market Assessment and Evaluation.* (CBEE Technical Services Consultants 1998).

For both market characterization and baseline measurement, the types of market indicators to be studied should be driven largely by the (revised/assessed) program theory. In general, any key market indicator that the program theory predicts will be changed by the program should be considered for inclusion in either a market characterization or a baseline study. However, markets are constantly in a state of change. Likewise, program theories can and do change over time to reflect added knowledge of how a market works, to reflect changes in how the market operates and to reflect changes in a program's thinking about how to influence a market. If these changes are significant, a market characterization or baseline study that considered only those market indicators addressed under the initial program theory may be outdated before it is even completed. Thus it is important that market characterization and baseline studies seek to include not only those market indicators that are important under the initial program theory, but also those that could become important later. It is often the evaluator's job to predict what indicators may be important in the future and include those metrics in the study.

Often one or more market indicators can be created for each hypothesized outcome in the logic model. Similarly, there can be a testable hypothesis and an associated metric(s) embedded within the relationship between different program results. That is, multiple program results may be influenced by the same activities resulting from a single relationship or event hypothesized in the program theory. The ability to chart the flow from intervention to outcomes to further outcomes and the interactions of outcomes compared to the program theory is an important part of the evaluation effort. Assessing each link can provide information that can separate problems with the theory associated with the causal effects (the basis of program design) from failure of the program to set events in motion to achieve those effects. Disaggregating "theory failure" from "program failure" is a necessary assessment in order to improve the program should the causal expectations not go as planned. (A more detailed discussion of theory failure versus program failure can be found in Weiss.²¹⁴) Without this assessment, it is difficult to know what needs to be changed and whether a change has a likelihood of succeeding.

Market indicators will likely be selected based upon the theoretical foundations within the evaluation team's program theory. Programs designed to speed the adoption and use of energy efficiency products may use research methods that focus on study approaches that measure the rate of customers moving down the product diffusion path.²¹⁵ These measurement approaches focus on the extent to which customers move along the diffusion path steps of technology awareness, persuasion, decision, implementation and

²¹⁴ *Evaluation: Methods for Studying Programs and Policies*. (* Weiss 1998).

²¹⁵ Rogers describes the stages of the innovation-decision process as knowledge, persuasion, decision, implementation, and confirmation. See Chapter 5 in *Diffusion of Innovations*, 4th Edition. (Rogers 1995). Others, such as Hassinger in "Stages in the Adoption Process," argue that need is required before someone will listen to messages that will provide knowledge about the alternative. (Hassinger 1959). Other authors have looked at the continuum as awareness, knowledge, adoption, and confirmation. There have also been other variations and suggestions that the process may be somewhat different depending on the product or service and entities involved.

confirmation.²¹⁶ In addition, evaluators can also measure the elements associated with purchase behavior, including the number of each of the different supply chain actors involved with the technology, store stocking behavior, and technology advertising/promotion behavior. A program focusing on eliminating specific market barriers will need to have measurement of these barriers and how they might be changing over time. Generally, the ultimate outcomes would include a market indicator measuring an analysis of efficient market share, i.e., the percentage of the measures or services sold that meet the energy efficiency criteria.

Orienting Theoretical Perspectives

There is a maxim that is popular in certain academic fields suggesting that there is nothing quite as practical as a good theory. The same is true for MT program evaluation: rooting market effects studies within a broad theoretical perspective regarding how markets work can help lend support to the program theory, improve the credibility of results, and provide important insights that might otherwise have been missed.

A wide range of such theoretical perspectives on markets is possible. However, one framework that has been used to help shape many past market evaluations in California and elsewhere is the market barriers paradigm first developed in the 1996 *Scoping Study*.²¹⁷ Under this paradigm, markets are viewed as being pervaded by certain recurring structural characteristics (market barriers) that tend to lead to gaps between the actual level of investment in energy efficiency and the level of investment that would be societally optimal. Thus, a key objective of market characterization, baseline, and market progress studies is to document these barriers and the extent to which programs reduce them.

Another theoretical perspective that has helped shape past market evaluations is diffusion of innovations theory. Briefly, the literature on diffusion of innovation provides a general model describing how products and ideas penetrate markets.²¹⁸ The theory includes elements that account for the structural, social and cultural environment in which decision-making occurs, the characteristics of firms and individuals that influence decision-making, the criteria that potential adopters use in evaluating the characteristics of products, and the characteristics of the product itself. MT evaluation studies that draw on diffusion of innovations theory tend to focus on these issues, and on the manner in which energy efficiency programs advance the diffusion of energy efficiency technologies.

²¹⁶ Hall and Jordan evaluated pre and post-program movement through the diffusion of innovation steps for several USDOE programs in *An Evaluation of Selected Technical Assistance Services Provided by the Federal Energy Management Program: Results of a 1999 Customer Survey*. (Jordan and Hall 2001); and *The 2001 FEMP Customer Survey: Study Report*. (Hall and Jordan 2001).

²¹⁷ *A Scoping Study on Energy-Efficiency Market Transformation by California Utility DSM Programs* (* Eto et al. 1996).

²¹⁸ *Diffusion of Innovations*, 4th edition. (Rogers 1995).

Other theoretical perspectives to help shape market effects studies have been proposed, including transaction flow analysis, communications theory, and network analysis.^{219, 220} In addition, there is a vibrant and growing literature on the demonstrated utility of the various theoretical perspectives that have been applied and on the potential compatibility of competing perspectives.²²¹ In fact, Kunkle and Lutzenhiser argue that to create better energy efficiency policies and programs, a broadening of the perspectives and the paradigms examined should occur in measurement and evaluation.²²²

At the end of the day, the authors believe that what is most important is not to choose from among the available theoretical perspectives on markets, but simply to recognize the critical role that any broad theoretical perspective on markets can play in underpinning market effects studies.

Baseline Measurement: Some Practical Considerations

In examining the program goals evaluators frequently learn that the program will “increase,” “change,” or “improve” something, but the program planners may not know where they are starting from – the baseline. This problem needs to be anticipated.

- If a baseline is not available, it should be determined.
- It is important to recognize that baselines are not static points, but slopes of expected change.
- In measuring baselines researchers need to keep in mind that many changes can occur without a program.
- Sometimes it takes an evaluation method or process to obtain a good estimate of where the baseline will move in the absence of the program.²²³

²¹⁹ See, for example, “Using Diffusion and Communications Theory to Expand Market Barrier Examination in MT Measurement.” (Megdal et al. 1999).

²²⁰ “Integrating Perspectives from Alternative Disciplines to Understand Market Transformation Policy in Energy Markets.” (Megdal 1998). Though they do not use the term, most of the market barriers discussed in the *Scoping Study* (* Eto et al. 1996) can be found in earlier literature discussions within Transaction Cost economics. One example of literature that looks at market operation and change from a network perspective includes Rosen’s work on word-of-mouth marketing, “buzz”: *The Anatomy of Buzz: How to Create Word-of-Mouth Marketing*. (Rosen 2000).

²²¹ See, for example “Why Can’t We All Just Get Along? A Reconciliation of Economic and Innovation Diffusion Perspectives of Market Transformation.” (Mast 1999); and “Does Talking About Barriers Just Get in the Way?” (Goldberg 2003).

²²² “Beyond Market Transformation: Some Perspectives on Energy Evaluation and Research and the Energy Efficiency Movement.” (* Kunkle and Lutzenhiser 1999).

²²³ One such process, though not the only potential one, is a Delphi process that uses market and program experts in a way to gather their opinions while not allowing any one or small group of individuals to overly influence the group outcome. See “Dressing the Priestess: Preparation for and Results of a Delphi Study for a Residential New Construction Program.” (Blake et al. 2003). Recognize, however, that a Delphi process is not so much a forecast as it is a method to put together opinions without undue influence from forceful individuals. As such, it is only as good as the process undertaken to obtain a balanced and unbiased panel (e.g., not an over-representation of personnel involved with programs versus those knowledgeable about the markets).

- Even experts need to know where we are today in order to forecast where we might be over time.

Further, evaluators sometimes hear: “But the baseline is too hard (or costly, or vague, or changeable) to measure.” It is possible that something is too costly to measure, and that other surrogate variables may need to be identified,²²⁴ but clearly if something is too hard to measure, too changeable to get a handle on, or too vague to define, the evaluation is in trouble and the program may be meaningless. If we can’t measure it in the baseline, how will anyone measure it as a program progress indicator or outcome variable?

Market Indicators and Measuring Market Progress

Chapter 6 on Impact Evaluation was concerned about estimating the program’s “impact” on energy and demand savings. Impact evaluations are generally thought of as measuring the primary effects of resource acquisition efforts. MT programs also generally have ultimate goals of energy and demand savings. However, as discussed above these are obtained more indirectly through interventions designed to change the market as described within the program theory. (See the *2001 Framework Study* for a much more thorough description and the theoretical foundations for these various types of program interventions.) The effects being measured from MT programs are, therefore, those program induced market changes and ultimate changes. There are several differences between measurement of market transformation and resource acquisition efforts. These involve the questions being posed, the groups examined in the research, and how effects are measured.²²⁵ One of the key differences is the need to identify and measure market indicators in order to measure market change, and then market effects (as the degree of market change(s) that are program induced).

The selection and timing of market indicators to be measured in a MT program evaluation is as much an art as a science. However, discussed below are a number of general principles suggested either by the literature on market effects evaluation²²⁶ or the authors’ experience.

Another method is the dynamic forecasting method provided in Chapter 7 of the *2001 Framework Study*. This method was then used in Wisconsin. A summary is available in “Dynamic Modeling of Market Effects & Spillover with Limited Information.” (Goldberg and Agnew 2003).

²²⁴ Avoid getting trapped into looking at variables that are easy to measure, but aren’t really related to the target variable – “looking for the keys where the light is good, instead of in the dark place where you dropped them.”

²²⁵ A comparison of the measurement issues in impact evaluation versus MT evaluation is presented in “Methods & Measurement Issues for a DSM Evaluation Versus a MT Market Assessment and Baseline Study.” (* Megdal et al. 1999).

²²⁶ The term “MT program evaluation” is used to cover all the types of studies needed: PT/LM assessment/refinement, Market Characterization, Baseline Studies, Market Progress, and Causality Assessment. Market effects measurement or evaluation generally is used to refer to the subset of Market Progress (with comparisons to Baseline Studies) and Causality Assessment.

The Need to Tell a Story

Perhaps the single most important tenet regarding the selection and timing of market indicators to be measured is the need for the evaluation to tell a coherent story about the chain of events that the program is intended to cause to occur. This point was made first by Herman et al.,²²⁷ was elaborated on in the *2001 Framework Study*, and has been discussed in other papers as well. In general, “a coherent story” means that the evaluation should systematically explore the extent to which the chain of causally linked events hypothesized by the program theory is actually occurring. Toward this end, it is important both to include in the evaluation those market indicators that play the most pivotal role in the program theory, and to ensure that the evaluation is looking for changes in these indicators roughly when predicted by the program theory, rather than sooner or later.²²⁸

The Need for Targeted Studies in the Early Years

There is a tension in market evaluation between the need for timely feedback on program performance and the relative slowness with which most market transformation initiatives change energy efficiency markets. It may be years before meaningful change can be observed in overall market indicators.²²⁹ Feedback may be needed much sooner than that, either to refine the program, or to decide whether it is working well enough to be continued. This tension can be partially resolved through the judicious combination of well-linked proximate and ultimate market indicators. However, it may still be difficult to obtain feedback on market effects early enough in a program’s life cycle.

One approach that is useful for dealing with this challenge is to attempt to isolate subsets of market actors who can be expected to show attitudinal and/or behavioral changes particularly early if the program is working as expected – for example, market actors who come into direct contact with the program, including, but not necessarily limited to, program participants.²³⁰

Another primary use of early evaluation results is for testing the program theory and providing feedback to program planners and implementers to allow them to revise the PT/LM and refine the program design. Blumstein et al. describe the need for MT pilot efforts and using a continuous program improvement process as part of an approach based upon theory-based market transformation program development. These authors

²²⁷ “Measuring Market Transformation: First You Need a Story ...” (Herman et al. 1997).

²²⁸ Of course, it is often useful to measure market indicators well before they are hypothesized as changing, in order to provide a baseline. However, our focus here is on when the evaluation anticipates and tests for a *change* in these indicators.

²²⁹ “Market Transformation: Expectations vs. Reality.” (Hall and Reed 1999).

²³⁰ Looking for behavioral effects on the part of program participants in the context of a market evaluation should not be confused with the measurement of participant energy savings in the context of a traditional impact evaluation. The key difference is that in the former case we are typically looking for relatively lasting behavioral changes that are tied to the longer-term market effects predicted by the program theory, while in the latter case we are interested mainly in near-term energy savings, and the participant behaviors that drive these savings.

suggests that this system of program planning, pilots, and evaluation is needed due to the lack of experience and understanding by some of the program designers and operators concerning how to transform markets (rather than just make resource acquisitions). Evaluation and market testing can be used to maximize program-associated learning and minimize program design and operational risk. These authors argue for the critical role that early MT research provides in helping program administrators and others understand market dynamics and the role of program intervention within a market.²³¹ Another example of the value of early MT evaluation is the program theory and evaluation work conducted on the California Nonresidential Standard Performance Contracting (SPC) program in 1999. This study demonstrates the importance of using a PT/LM process and early evaluation findings as input for significant program improvement recommendations.²³²

The Usefulness of a Consistent Framework and Institutional Process

Hastie et al. describe a framework and institutional process used to drive the evaluation of market transformation initiatives at the Long Island Power Authority (LIPA).²³³ Under this framework, program designers complete a table using a standardized format, showing the chain of specific behavioral effects the program is expected to have on market actors. Columns of the table show the categories of market actors influenced, rows show different time periods ranging from less than one year after program implementation to more than five years, and individual cells show the specific behavioral change anticipated for a given market actor in a given period. The table is accompanied by a narrative summarizing the causal links between all of the anticipated behavioral changes. Program evaluators critique the table and narrative developed for each program as part of the development process. Following the review by the evaluators, the program designers revise it as appropriate. The resulting “standardized” program theory is then used to drive the selection and timing of market indicators to be measured in evaluating the program.

Hastie et al. conclude that this framework had a number of advantages, including: (1) illustrating deficiencies in the logic or understanding of specific market segments; (2) illuminating linkages between stimuli and effects and among market effects; (3) making the market transformation approach more tangible to the utility’s Board of Trustees; (4) making obvious the market indicators that needed to be tracked; and (5) serving to ensure that all key market effects were being targeted by specific implementation activities. LIPA’s experience would seem to suggest that it might be beneficial for California to develop a consistent framework and institutional process for the selection of market indicators to be studied in market transformation evaluations.

²³¹ “A theory-based approach to market transformation.” (Blumstein et al. 2000).

²³² “Applying a Theory-Based Approach to California’s Non-Residential Standard Performance Contract Program: Lessons Learned.” (Goldstone et al. 2000).

²³³ “A Systematic Application of Theory-Based Implementation and Evaluation of Market Transformation Programs.” (* Hastie et al. 2000).

Dealing with Large Numbers of Market Indicators

A number of market indicators may be desired to be measured either as part of assessing the various program logic model elements and assumptions or as part of tracking market changes. One method of assessing a large number of market indicators is through the use of a binomial test, as proposed by Shel Feldman and used in a market evaluation by a GDS Associates team for the Boston Gas Company.²³⁴

Don't Forget About Gross Unit Savings

One of the distinguishing features of market transformation program evaluation is that it attempts to measure things occurring in the market in order to test the program assumptions. Social scientists have developed methods to measure many program or program-related effects, but in energy efficiency program evaluation, the ultimate outcome metrics about which policy makers care the most are energy and demand savings. While it is important to measure intermediate steps to claiming market effects, the ultimate value will be based on the energy effects. Identifying a program goal as a “market effect” doesn’t mean that the laws of physics are suspended. Evaluators and planners still need the same basic foundation of savings per widget that is required for resource acquisition programs – the only difference is that for market transformation programs, estimates are derived by applying savings metrics to an estimated market change, rather than a building, a set of buildings, or some other easily identified physical unit (Principle 3).²³⁵

²³⁴ For a summary of this work see “Measuring Market Transformation Progress & the Binomial Test: Recent Experience at Boston Gas Company.” (* Spellman et al. 2000).

²³⁵ It is worth noting, however, that there may be differences between the ideal methods for estimating gross unit savings for a market transformation initiative and a resource acquisition program. In the former case, the uncertainties surrounding gross unit savings may be swamped by those surrounding the number of measures the program caused to be adopted. Thus it may make sense to focus the bulk of available evaluation dollars on establishing attribution.

Causality Assessment²³⁶

It is important to recognize that changes observed in the market are not necessarily market effects induced by the program.²³⁷ Market effects were defined in the *Scoping Study* to be explicitly different from market changes by this fact.

*A market effect is “a change in the structure of a market or the behavior of participants in a market that is reflective of an increase in the adoption of energy efficient products, services, or practices and is causally related to market intervention(s).”*²³⁸

The fourth principle for a high quality MT evaluation is to critically assess causality and sustainability. The most difficult of the key evaluation planning elements is setting up the evaluation to approximate causal arguments. When examining markets, it is seldom possible to prove causality. Markets are complex and dynamic. Designing programs that contain experimental design through implementation might be ideal. Yet, this is seldom possible and market operations can interrupt the intended experimental design. Evaluators often then can provide evidence of causality, but not proof.²³⁹ Some of the better ways of building evidence include:

- An *a priori* description of the process and steps to a changed market condition that can be traced in the evaluation.²⁴⁰

²³⁶ An early cautionary note regarding this term seems appropriate. There are at least two terms in relatively widespread use to describe the process of assessing the extent to which an energy efficiency program has played a causal role in observed improvements in market functioning: “causality assessment,” and “attribution.” Both terms have their merits, but both come with pitfalls. On the plus side, the term “causality assessment” has a venerable history in the social sciences, as well as in the evaluation of other kinds of social programs, suggesting that this term offers the advantage of recognizing that energy efficiency program evaluation does not exist in an intellectual vacuum. On the minus side, the term “causality assessment” carries with it a certain hubris, given the profound difficulties of reliably assigning causes for social outcomes, some of which are discussed in this section. On the plus side, the term “attribution” seems more appropriately humble in its connotations. However, on the minus side, this term often seems to end up getting grossly misused in practice – as, for example, when it is used to refer to purely administrative negotiations over which of two overlapping programs should get to take credit for an observed market effect. On balance, the term “causality assessment” has been selected, but the provisional nature of all conclusions regarding the causes of market changes needs to be emphasized.

²³⁷ See *Market Effects Summary Study, Volume 1* (of 3). (Peters et al. 1998), page 33.

²³⁸ *A Scoping Study on Energy-Efficiency Market Transformation by California Utility DSM Programs* (* Eto et al. 1996), page 9.

²³⁹ To understand the difficulty of the task, one must consider that even true experimental designs with large samples, and ones which are replicated several times, may not show a causal effect, but merely a consistent correlation. An experimental design that adds and removes a stimulus under laboratory conditions can provide both necessary and sufficient evidence that the cause and effect are identified – but this is often not an option available to public purposes programs.

²⁴⁰ There is significant literature within program evaluation on this method. For example, see two proponent articles and one opposed position article in “Program Theory in Evaluation: Challenges and Opportunities.” (* Rogers et al. 2000b); “Ascertaining Causality in Theory-Based Evaluation” (Davidson 2000), “Program Theory in Evaluation: Causal Models in Program Theory Evaluation”

- A quasi-experimental design where some, but not all extraneous, confounding factors can be controlled.
- A “comparison group” design (as opposed to a pure, randomly assigned control group) that allows the researchers to find a similar market without the program impetus, but hopefully with other important factors at work.
- An explicit exercise in which reasonable alternative hypotheses are considered and shown to be less supported by the facts.
- Interviews with market actors, who may provide convergent evidence because the market actors in a program may have a clear picture of why consumers took action.

An early MT evaluation looked at the effects of lighting efficiency programs in New England and assessed causality by using the reduction in market barriers as part of the evidence for attributing the changes seen in the market as being caused by the program. This work is summarized in Meber et al.²⁴¹ This study is cited in the *2001 Framework Study* with the comment that this work notes that “the real problem is causality and whether any other factors could account for the observed changes” (pages 5-22).

An example of the use of quasi-experimental design, pre-post analyses, and comparison group analyses for a causality assessment for a market transformation program is provided by an analysis done for one of the New York State Energy Research and Development Authority’s (NYSERDA) programs. This causality work is summarized in a paper by Megdal et al.²⁴² This work is reported in NYSERDA’s 2003 Evaluation and Status Report and was continued in its 2003 Evaluation and Status Report. (Both are available at <<http://www.nyserda.org/publications.html>>.)

Causal linkages between training of sales personnel in retail lighting and appliance stores with changes in ENERGY STAR[®] equipment was tested in California and reported in Ridge et al.²⁴³ The study found evidence to support some but not all the causal linkages tested. Alternative explanations and external influences were assessed and discussed.

After a design is chosen that will put together an argument for causality (attribution) – if the research supports it – the question of measurement of savings still needs to be answered. Market effects can usually be measured with confidence (assuming the baseline, measurability, and causality elements have been met), but frequently without much precision. Measuring changes for direct participants and tight definitions of participant spillover are easier to make precise. It is much more difficult to measure exact program induced energy savings when market effects involve a several step chain of causal logic.

(* Rogers et al. 2000a); and the opponent piece: “The False Choice Between Theory-Based Evaluation and Experimentation.” (Cook 2000).

²⁴¹ “Converging On the Effects of Utility Lighting Efficiency Programs.” (Meber et al. 1997).

²⁴² “He Did It! He Did It! – Providing Evidence for Causality.” (* Megdal et al. 2001).

²⁴³ “Testing the Causal Linkage Between Training of Sales Personnel in Retail Lighting and Appliance Stores and Changes in Market Share of ENERGY STAR[®]—Qualifying Equipment.” (* Ridge et al. 2001).

For example, Feldman and colleagues provided significant evidence supporting the Consortium for Energy Efficiency's collaborative effort as a major contributor to the transformation of the residential clothes washer market.²⁴⁴ In the Pacific Northwest, ENERGY STAR® windows started with a baseline penetration of about fifteen percent of windows sold. Three years later, over sixty percent of the windows sold were ENERGY STAR. This did not happen anywhere else in the country (comparison groups), almost no utilities paid incentives (alternative hypothesis), and the window manufacturers unanimously credited the program in making this happen (knowledgeable market actors). The evidence and the confidence that these programs significantly contributed to these changes is overwhelming, but precisely how many gigawatt-hours were saved is not easy to know. In this case, although the effects are large and clear, policy makers may need to accept somewhat imprecise estimates.²⁴⁵

Even with a clutter of multiple programs and historical changes confounding the market baseline, it is still possible to ferret out clear evidence of market effects attributable to a planned intervention. An evaluation of the CFL program in California during the energy crisis by Rasmussen and her colleagues conducted such an exam.²⁴⁶

An alternative finding is seen in an evaluation for the "Just Enough Air" program. This program attempted to get wood shops to reduce the vacuuming force of their air-sucking equipment. The results showed that the baseline conditions were higher than expected, the market was smaller than thought, most of the participants said they wouldn't pay for it without utility incentives, there were very few additional market players after the program, and the only spillover was what might happen as a result of disseminating the "how-to" manual that the program produced. There did not appear to be near-term or long-term effects attributable to the program.²⁴⁷

Market effects are complex and often difficult to achieve, but when they are achievable, they should also be measurable. The challenges to evaluating market effects are real, but they are not overwhelming. Evaluators have more to track when looking at market effects than when looking for direct effects of acquisition programs. The program logic model may be complex, and many market activities and relationships may have to be measured and compared in order to meet the challenges of causality. This may obscure the need to continue to obtain the measurements required to estimate demand and energy

²⁴⁴ *The Residential Clothes Washer Initiative: A Case Study of the Contributions of a Collaborative Effort to Transform a Market.* (Shel Feldman Management Consulting et al. 2001).

²⁴⁵ In the northwestern windows example, the baseline could have grown to 18% or 24% in the absence of the program. The forward strength of the momentum after the utility support ended may add 15 – 25% more penetration without further intervention. Frankly, the actual savings from changing the rated U-value of the square foot of window area will probably never match the engineering estimates. An approximate estimate of savings can be made but it will not be precise. (Quantec 2002).

²⁴⁶ "Addressing Program Attribution in the Wake of the California Energy Crisis." (Rasmussen et al. 2003).

²⁴⁷ *Just Enough Air: Efficient Pneumatic Conveying. Concluding Memorandum.* (Stout and Scott 2003); and *Fan Speed Reduction in Pneumatic Conveying Systems in Secondary Wood Products Industry.* (Quantec 1999).

savings. It is not useful when accounting for public money to say that the market effects are large, but that savings can't be reasonably estimated. Demand savings will follow from the end use load shape of the relevant measures. The assumptions behind the savings estimates for each widget should be tested when the market effects are counted on to avoid power purchases.²⁴⁸

The challenge of identifying “net” savings of market effects is not much greater than for direct acquisition programs. If the baseline is known with confidence, calculating the net savings should be very easy. As it is the baseline of naturally occurring or non-program induced savings is an *estimate* that is made less certain by the fact that it must be *forecast* into future years in a hypothetical situation. However, every method for getting at net savings for direct acquisition programs is confounded by uncertainty (see Chapter 6 on Impact Evaluation). In addition to self-report problems, statistical estimation issues, and incomplete information, none of the field-tested methods assess the full impact of the non-participant spillover. When information spillover is large, many of the actions taken by non-participants may be a result of the program or the program portfolio. Likewise, many of the actions taken by free riders, may, in fact, be actions that are taken as a result of information spillover from the program or the program portfolio when that information is spread through the market.

Programs need not be justified by their intended market effects, but if these effects are important, they can and should be evaluated and documented.

Assess Sustainability and Plan to Measure Long-Term Market Effects and Post-Program Sustainability

There are two fundamental challenges surrounding assessment of the sustainability of market effects. The first is that sustainability is, by definition, a long-term concept. With energy efficiency markets, as with any human institution, the more time that passes, the greater and more complex the variety of social forces that come to bear, and thus the harder it is to reliably and coherently assess the reasons underlying observed changes. The second fundamental challenge is that, again by definition, sustainability focuses on what happens, or would happen, to the market after a program is withdrawn - and more often than not an assessment of this issue needs to be made while the program is still in operation.^{249, 250}

²⁴⁸ In a forthcoming report on the evaluation of the effectiveness of the Northwest Energy Efficiency Alliance, the evaluators suggest that the Alliance continuously test the underlying energy savings estimates for their programs. When these savings are multiplied, in order to estimate the long-term savings from the market effects, the resulting calculations need to be based upon the most accurate impact estimates.

²⁴⁹ Indeed, to withdraw a market transformation initiative before there is good evaluation evidence regarding the sustainability of observed market changes would be imprudent, as it would unnecessarily risk premature loss of the observed program benefits.

²⁵⁰ See “Saturation, Penetration, Transformation: How Do You Know When a Market Has Changed?” (* Grover et al. 2002) for a study that begins to assess large changes in penetration versus what may be sustainable as market transformation.

The combination of these two challenges suggests that assessment of sustainability relies as much on reasoning as on empirical evidence. For example, the following are some bases for asserting sustainability that have been suggested by industry observers.²⁵¹

- *Relative Irreversibility*, or the existence of demonstrated market effects that are sufficiently fundamental that they would be difficult to reverse once the program is removed. An example of relative irreversibility that is often used is the retooling of manufacturer production lines to favor more efficient technologies in a manner that would be expensive to change back.
- *Fundamental Changes in the Incentives Facing Market Actors*, which may occur when a market transformation initiative causes beneficial changes in the basic market signals facing one or more categories of market actors. This may involve market structure changes that eliminate the market barriers or the development of profitable market entities that continue to drive market transformation. For example, many residential new construction programs over the past several years have sought to increase homebuyer awareness of energy efficiency, reasoning that significant increases in awareness can be expected to lead to more builders positioning themselves as energy efficient, and once a builder has successfully positioned itself this way, it is in its interest to continue to capitalize on the reputation it has established.
- *New Codes and Standards*, which use the force of law to institutionalize program-induced gains in energy efficiency.
- *Disappearance of Inefficient Technologies from the Market*. An example would be magnetic lighting ballasts for commercial facilities, which, following significant program-induced acceleration of the diffusion of electronic ballasts, have come to be relatively rare, at least in larger buildings.

What all these potential foundations for asserting sustainability have in common is that they require both a detailed understanding of the specific manner in which a program has changed the structure or functioning of a market (an empirical component) and a systematic argument as to why either human nature, economic forces, or laws can be expected to preclude these changes from reversing themselves (a reasoning component).

Despite the fact that an initial assessment of sustainability would generally be made while a program is still operating, it is difficult to overstate the desirability of continuing to systematically monitor energy efficiency markets once they have been deemed transformed and the energy efficiency programs have been withdrawn. All too often, such post-program tracking falls by the wayside, as markets that are still being actively targeted for intervention absorb the bulk of evaluation resources. In those relatively few

²⁵¹ See, for examples: “The Elements of Sustainability.” (Hewitt 2000); *Market Effects Summary Study, Volume 1* (of 3). (Peters et al. 1998); a summary can be seen “Measuring Market Transformation: The 1997/1998 California Market Effects Studies.” (* Peters et al. 1998); “Measuring the Market Effects of Utility Programs: Lessons from California.” (Mast et al. 1998); and *CTAC Market Effects Study* (Hagler Bailly Consulting 1998).

cases in which such post-program monitoring has been done for an extended period of time, the body of evidence regarding sustainability has been mixed. For example, in reviewing evidence from Wisconsin, Prah and Pigg²⁵² found that in some cases market effects appeared to persist, while in other cases markets tended to partially regress toward their pre-program state.

Finally, once a reasonable case has been made for sustainability, a forecast of the long-term effects of the program on measure adoption is in order to support cost-benefit analysis. Methods for doing so are a frontier area in MT program evaluation. To date, most efforts to systematically forecast long-term program-induced acceleration of measure adoption have relied on Delphi analysis, under which a panel of experts provides individual forecasts, information on these forecasts is returned to the panel without being attributed to individual panel members, and panel members use this information to revise their forecasts.²⁵³ The *2001 Framework Study* provides an extensive discussion of alternative approaches, including a dynamic modeling approach that relies on modeling the adoption process for repeated time periods. To our knowledge, most of this proposed approach has yet to be implemented anywhere. Recently, however, a pioneering effort in Wisconsin has used a type of dynamic modeling approach to forecast the long-term market effects of a set of programs targeting business customers.²⁵⁴

Needless to say, given the earlier discussion of the challenges inherent in assigning causality for even near-term market effects, it is important to remember that even the most rigorous methods for modeling long-term market effects can be regarded as producing highly uncertain results.²⁵⁵ Given this, when the time comes to use the results of such modeling in a cost-benefit analysis, it may be appropriate to adopt non-traditional approaches that explicitly acknowledge the presence of extensive uncertainty – for example, analyzing program benefits under multiple scenarios, or focusing on assessing the *probability* that benefits will exceed costs, rather than the more common focus on developing a single point estimate for the ratio of benefits to costs.

²⁵² “Do the Market Effects of Utility Energy Efficiency Programs Last? Evidence From Wisconsin.” (Prah and Pigg 1997).

²⁵³ For an example of Delphi analysis in practice see “Dressing the Priestess: Preparation for and Results of a Delphi Study for a Residential New Construction Program.” (Blake et al. 2003). A word of caution here. The “forecasts” are the informed opinions of individuals knowledgeable about the industry, but there are not formal forecasts that statistically are formed from extensive data analysis and modeling. Given this, it is extremely important that the Delphi panel be selected to be balanced between program supporters and those only knowledgeable about the industry. Otherwise, a biased forecast could easily be expected. Similarly, the information provided to the panel and in what form could also create bias. So great care and expertise is needed to properly use this technique and minimize potential bias.

²⁵⁴ *Business Programs Evaluation: Market Effects Pro Forma Estimates*. (Goldberg et al. 2003); and “Dynamic Modeling of Market Effects & Spillover with Limited Information.” (Goldberg and Agnew 2003).

²⁵⁵ This does not mean that market transformation should be viewed as a second-class resource. All efforts at long-term forecasting must deal with considerable uncertainty – including forecasting for both energy efficiency resource acquisition, and supply-side construction planning, two of the main alternatives to market transformation.

Chapter 11: Non-Energy Effects Evaluation

Preface

This chapter of the Framework presents the non-energy effects roadmap and issues surrounding non-energy effects evaluations of California's energy efficiency programs. Non-energy effects (NEEs) are those net effects of a program other than energy related (energy or demand) impacts. These might include reduced water use, improved environmental conditions, higher productivity, or other similar benefits. However, these effects may also include increased water use, lower productivity, or other changes that have a negative impact on the participant.

This chapter presents a definition for NEEs evaluations, a discussion of the reasons why NEEs evaluations are included in the Framework, and a discussion of the skills needed to conduct NEEs evaluations. The chapter also provides a discussion of the typical evaluation tools and evaluation approaches for researching NEEs, and examples of the various types of NEEs evaluations. The chapter ends with a presentation of the decision steps associated with planning and conducting NEEs evaluations within the Framework.

Introduction and Key Issues

This chapter focuses on the evaluation of non-energy effects within the Framework and specifically identifies the conditions under which Public Goods Charge funded programs can spend evaluation resources to conduct NEEs evaluations. These conditions are specific to the State of California and are grounded within a set of policy decisions that limit the conducting of NEEs research to effects that are approved for research by the CPUC, or on effects in which a specific program's success at obtaining energy resources is directly influenced. This policy direction was provided to the authors by the CPUC staff during the development of the Framework.

An example of a program that could justify NEE evaluation research might be one that promotes energy efficient clothes washers. This type of program not only emphasizes energy savings, but also focuses on reducing the amount of water used. The amount of water saved is one of the key customer effects that cause or influence participation levels. In this case, the level of water savings is a direct determining factor relating to the amount of energy saved. For this evaluation, resources could be spent documenting the amount of water saved in order to confirm the presence of the benefit, and to use those results in promotional efforts to increase participation. Another example would be a case in which a CPUC policy decision establishes a need to evaluate the level and type of generation-associated emissions reduced as a result of California's energy programs. In this case, evaluation resources could be spent on quantifying the level of emission reductions and the value of those reductions. The Framework's NEEs research decision steps presented at the end of this chapter provide additional guidance on this topic. These

decisions significantly limit the NEEs research that might be conducted in California for the types of programs covered by this Framework.²⁵⁶

While in some circumstances NEEs evaluations can be conducted under the Framework, the results from those evaluations are not to be included in cost-effectiveness tests used to evaluate the program's ability to save energy (kWh or therms) or to influence demand. However, the results of the NEEs evaluations can be used to help define the total public value of the programs offered (such as in a public purpose test).

Definition

For the purposes of this Framework, non-energy effects are defined as *any program implementation or participation effect that is other than the direct energy (kW, kWh, therms) effect(s) associated with an energy efficiency, resource acquisition, or resource procurement program funded in California through the use of Public Goods Charge or procurement funds that falls within the types of programs covered under this Framework.*

This is, by design, an encompassing definition that includes a wide range of effects. These NEEs can include effects on the program participant, such as increased or decreased levels of productivity, sales, water usage, or comfort, to name a few. Non-energy effects can also be effects on the society as a whole, such as lower or higher levels of emissions from generation facilities and the associated changes in the health and welfare of the populations living within the areas in which these emissions are released. Likewise, NEEs would also include changes in the need for water supply and treatment facilities as a result of changes in water consumption levels resulting from program-installed measures. There may also be NEEs on the companies selling or marketing energy that are a result of the actions taken through an energy efficiency program, including changes in bill collection costs, levels of debt owed by customers, or company productivity or profitability.

In presenting this definition, the reader will notice that NEEs are defined not as "benefits" (as typical in other literature), but as conditions that can change for the benefit or the detriment of the participant, society, or the energy provider. This definition is used because NEEs research has documented both negative and positive effects relative to one or more program outcomes.²⁵⁷ While the vast majority of NEEs are positive, and provide added value, there are cases in which the effects are negative. In order to be objective, it is important for this definition to recognize that NEEs can be positive or negative. This definition fills that requirement.

²⁵⁶ It should be noted that this Framework does not influence the research conducted for California's low-income programs. These programs have their own evaluation decision system that is beyond the jurisdiction of this Framework and which allows for a broader focus on non-energy effects research.

²⁵⁷ *Non-Energy Benefits to Implementing Partners from the Wisconsin Focus on Energy Program* (* Hall and Roth 2003).

Why Non-Energy Effects are Included in the Framework

The evaluation and reporting of energy program-induced NEEs is one of the fastest growing fields of energy program evaluation research. The primary reasons for this growth are:

1. The movement toward Public Goods Charge funded energy programs, in which energy consumers pay a small part of their utility bill towards a “Public Goods Fund” has increased the need to document the range of effects the public receives in exchange for those dollars. The central questions being asked by policy makers are: What is the total value the public is receiving in exchange for their Public Goods Charge dollar? In addition to energy savings, are there other societal benefits that are being achieved with these funds? Because energy programs may provide more than energy savings, public policy and public value accounting requirements suggest that there is a need, where appropriate, to document a wider range of effects the public receives in exchange for their program contribution.
2. Several evaluation studies have concluded that participants in energy programs often implement efficiency measures for reasons other than saving energy. For many participants, the energy savings are not seen as significant enough to make the change, but instead contribute to the value of the total package associated with the change. In many cases, the value of the NEEs to the customer can often be greater than the value of the energy savings. Customers may view their energy savings as a lower-priority byproduct of the change, rather than the primary reason for the change. For example, a California study of schools found that students in well-designed energy efficient daylit classrooms had substantially higher test scores than students in classrooms illuminated by artificial light.²⁵⁸ The value of the increased learning to the parents and to the government body responsible for education can be priceless, just as the value of the increased learning to the child contributes to a potentially more productive life. Another California study found that retail sales increased significantly in energy efficient stores lit by natural light.²⁵⁹ In this study, the energy efficient lighting system led to a 30 to 40 percent increase in sales (controlling for other variables), dwarfing the effects of the energy savings. In another study on the effects of weatherization measures, researchers found that participants reported that they were more comfortable and that the increased level of perceived comfort was significantly more valuable to them than the energy savings.²⁶⁰
3. Program managers can focus their marketing and outreach efforts on promoting the NEEs of a program in order to motivate customers to participate in their programs or to take actions. As a result, evaluators are being asked to provide

²⁵⁸ *Daylighting in Schools: An Investigation into the Relationship Between Daylighting and Human Performance* (Heschong Mahone Group 1999).

²⁵⁹ *Skylighting and Retail Sale: An Investigation into the Relationship Between Daylighting and Human Performance*. (* Heschong Mahone Group 1999)

²⁶⁰ California Low Income Public Purpose Test. (TecMRKT Works 2001).

additional information on NEEs for use in the program's marketing and sales efforts. Program personnel that do not understand the range of effects associated with energy programs are at a disadvantage when approaching a potential participant. Non-energy effects research can help improve the effectiveness of energy programs by demonstrating a wider range of benefits that can be used to increase participation and build program impacts.

4. Increasing environmental awareness of the general public, state legislatures, and the regulatory community have motivated people to look beyond the direct effects of programs and to investigate the indirect effects of energy programs. In particular, there has been a renewed interest in promoting energy efficiency programs as a solution to local and regional air quality problems and as a way to reduce greenhouse gas emissions.²⁶¹

These four reasons have focused considerable attention in some jurisdictions on evaluating the NEEs associated with energy programs. Taken together, the need to understand the full range of NEEs in the commercial, industrial, institutional, and residential markets can be important to regulators and organizations implementing Public Goods Charge funded programs. It can also be beneficial to program managers and sales staff who need to "sell" their programs.

Accordingly, there may be a need to be able to evaluate these effects under the Framework.

Evaluation Timing

Non-energy effects evaluations can be conducted at any time within the program design and implementation cycle, and in many cases after the cycle. However, there are some key considerations for the timing of these evaluations associated with the Framework. These include:

1. If the evaluation is being conducted to help support program enrollment and participation goals, the research should be conducted in time that the results can be used by the program manager to increase enrollments or installed measures. The results should be provided in time for program marketing materials to be developed, for staff to be trained on the use of the information, and for the effective use of the information to increase enrollments. Accordingly, the NEEs evaluation may need to be conducted in the early months of the program, but after enough participation has accumulated to conduct the evaluation. For example, an evaluation of an energy efficiency program that saves water heating costs by reducing hot water use will need reliable information on the amount of water and water cost savings the program provides in order to use these data in program outreach materials.

²⁶¹ *Scenarios for a Clean Energy Future*. (Inter-Laboratory Working Group 2000), and "Opportunities for promoting energy efficiency in buildings as an air quality compliance approach." (Vine 2003).

2. If the evaluation is conducted to support a CPUC policy to better understand NEEs, then the evaluation will need to be conducted before the date on which the CPUC needs the information. This timeline may, or may not need to be coordinated with the program's implementation cycle. For example, if the CPUC issued a policy directive that the Commission needs to understand whether California waste streams are influenced by the energy efficiency changes to industrial operations, the study would need to identify an appropriate timeframe in which to assess the changes in waste steam levels relative to the programs that influence those levels. This effort would require coordination with the program implementation timelines. However, if the CPUC needs to quantify the greenhouse gas reductions of these same programs, then an evaluation of the emissions reductions may not need to be conducted during the program cycle. This type of an evaluation may want to conduct a time dependant valuation²⁶² of the emission levels associated with energy production and use within different parts of the State and then model the time dependant load effects of those programs.²⁶³

Information on the NEEs needing to be quantified for a particular program may also be available from other sources. For example, a NEEs evaluation may use information (i.e. contact information) from other programs or program activities that best match the effect(s) that need to be researched. As an example, if program managers wanted to design information materials on the value of improved levels of comfort associated with a residential program, the evaluators may want to interview participants from past programs that provided similar types of measures to similar types of customers.

The above examples show that the timing of specific NEEs research may need to be tailored to the specific effects that need to be quantified or to meet specific policy requirements. There is no simple rule concerning when NEEs research should be conducted. Rather, each project needs to set a research timeline to match the effect being measured, the relationship between the effect and the program implementation efforts, and the needs of the information consumer.

Skills Needed to Conduct Non-Energy Effects Research

The non-energy effects being researched and the evaluation approach considerably influence the skills needed to conduct the NEEs research. These studies can require an assortment of skills available within the evaluation community and can cut across the skills normally associated with individual firms or evaluation professionals. While there is no general rule that can be applied to the skills needed to conduct NEEs research, a range of the skills and experiences associated with NEEs research include:

- Expertise in designing objective research approaches that match the information available to the researcher and the research budgets,

²⁶² *Time Dependent Valuation (TDV) Economics Methodology*. (Heschong Mahone Group 2002).

²⁶³ "Wisconsin's Public Benefits Approach to Quantifying Environmental Benefits: Creating Different Emissions Factors for Peak/Off-Peak Energy Savings," (Sumi et al. 2002a).

- Understanding of and experience with the application of experimental and quasi-experimental research designs,
- Expertise in data analysis methods, software, and analysis approaches,
- Experience in hypothesis and alternative hypothesis testing for causal relationships or indicators of causal relationships,
- Experience in standard evaluation research techniques, including:
 - Survey design and administration techniques,
 - In-depth interview design and implementation,
 - Secondary data mining and analysis.
- Experience in reporting and presenting evaluation research findings, including:
 - Ability to develop clear evaluation reports that describe the evaluation approach, the reasons for the selected approach, the strengths and weaknesses of the approach, the analysis conducted in enough detail that it can be replicated, and a presentation of the evaluation results.
- Experience in and an understanding of the subject area being studied (e.g., indoor air quality, productivity, greenhouse gas emissions, etc.):

Evaluation Planning and Approach Tools

All NEEs evaluations start with a detailed evaluation plan that presents the researchable issue(s) being investigated, the evaluation approaches considered, the evaluation approach selected, and the strengths and weaknesses of the selected approach. The plan should also include details of the analysis to be done, including the data to be obtained and analyzed, sampling approaches and sample size selection criteria, levels of precision and confidence intervals, and a presentation of the anticipated levels of uncertainty associated with the research approach and the data assessment efforts to be employed.

There is not a standard approach for quantifying the levels of NEEs associated with energy programs. These studies examine a wide range of potential effects and require flexibility in the evaluation approach. The NEEs evaluation may employ one or more of the following research methods as appropriate to match the research goals:

- Surveys: telephone, mail, Internet, e-mail,
- Interviews: in-depth, small group,
- Focus groups,
- Site visits and effects verification visits,
- On-site metering or monitoring of effects or anticipated effects,
- Secondary data acquisition, review, and analysis,
- Billing and payment data analysis,
- Economic or impact modeling tied to primary data collection or secondary data analysis,
- Pre- & post-program participation measurements and assessments,
- The use of comparison groups so that the net effects of the program can be estimated, and
- Other methods as appropriate for the effects being studied.

Research Methods and Approaches

As indicated above, there are many approaches that can be used to conduct evaluations of NEEs. This section of the non-energy effects chapter presents a small sample of the research approaches presented in the literature that have been used to estimate the presence or absence of specific NEEs, and in most cases, report their estimated value. This presentation is not meant to be exhaustive, but illustrative of the range and type of approaches used in this field. The examples presented in Table 11.1 include the results of low-income, residential, and commercial and industrial program evaluations in order to provide a range of approach examples.

Table 11.1: Examples of Non-Energy Effects and Research Designs

Program	Effect investigated	Methodology used	Reference
Wisconsin's Public Benefits Charge Funded Business Programs	Presence and value of increased productivity from installed energy projects.	In-depth interviews with participating partners who installed measures incentivized by the program.	Hall and Roth ²⁶⁴
Several Different Programs and Case Studies	Increased productivity in the industrial sector from energy efficient improvements.	Economic modeling of the results of several case studies and impact reports with productivity assessments.	Worrell et al. ²⁶⁵
Ohio's Weatherization Program	Value of reduced emissions from weatherization measures.	Application of traded values for SO ₂ , and application of NO _x control costs for burners and over-fire technologies to the estimated levels of reduced emissions.	Hill et al. ²⁶⁶
Detroit Edison's low-income energy education program	Actions and behaviors taken as a result of energy program training.	Survey of participants 6 months to 1 year after participation compared to a matched control group.	Hall and Reed ²⁶⁷

²⁶⁴ *Non-Energy Benefits to Implementing Partners from the Wisconsin Focus on Energy Program.* (* Hall and Roth 2003).

²⁶⁵ *Productivity Benefits of Industrial Energy Efficiency Measures.* (Worrell et al. 2001).

²⁶⁶ "The Environmental Benefits of Low-Income Weatherization," (Hill et al. 1999).

²⁶⁷ *Residential Low-Income Energy Management Program: Process Evaluation Report.* (Hall and Reed 1997b).

Table 11.1: Continued

Program	Effect investigated	Methodology used	Reference
Iowa's State Weatherization Program	Increased local area economic activity as a result of program spending.	Economic "input - output" model of economic effects.	Pigg ²⁶⁸
Residential Energy Efficiency Programs	Influence of energy bills on real estate value.	Regression analysis of American Housing Survey Data on housing value and energy consumption. In-depth interviews with high efficiency homebuilders.	Nevin and Watson, ²⁶⁹ Hall and Riggert ²⁷⁰
Missouri Gas Energy's, Washington State, and Oregon State Weatherization Programs	Reduced arrearage levels for program participants.	Pre and post program billing data analysis of arrearage levels and payment data using matched comparison groups of non-participants.	Hall and Reed, ²⁷¹ Khawaja et al. ²⁷²
Colorado Public Service's Energy Savings Partners Program	Reduced number of emergency calls as a result of home weatherization.	Uses the difference in the frequency of emergency calls to the utility between participants and non-participants.	Magourik ²⁷³
Detroit Edison's Residential Energy Management Program	Changes in the number of bill payments made on time as a result of program participation.	Analysis of billing and payment data for program participants before and after program participation.	RLW Analytics ²⁷⁴
U.S. Army's Energy Efficient Military Barracks Construction	Incidents of increased respiratory disease in newer, more energy efficient housing.	Comparison of medical records of Army recruits living in older, leaky, uninsulated barracks with those living in newer energy efficient insulated barracks.	Knoppel and Wolkoff ²⁷⁵

²⁶⁸ *An Evaluation of Iowa's Low-Income Weatherization Efforts.* (Pigg 1994).

²⁶⁹ "Evidence of Rational Market Valuation for Home Energy Efficiency." (Nevin and Watson 1998).

²⁷⁰ *Non-Energy Benefits Cross Cutting Report: Year 1 Efforts.* (* Hall and Riggert 2003).

²⁷¹ *Process and Impact Evaluation of Missouri Gas Energy's Pilot Weatherization Program.* (Hall and Reed 1998).

²⁷² "Effects of Weatherization Programs on Low-Income Customer Arrearage." (Khawaja et al. 1992).

²⁷³ "Evaluation of Non-Energy Benefits from the Energy Savings Partners Program." (Magourik 1995).

²⁷⁴ *Detroit Edison's 1995 Residential Energy Management Program Arrearage Analysis.* (RLW Analytics 1997).

²⁷⁵ "Chemical, Microbiological, Health and Comfort Aspects of Indoor Air Quality – State of the Art." (Knoppel and Wolkoff 1992).

Partial Listing of Non-Energy Effects Presented in the Literature

The NEEs literature includes a significant list of documented and anticipated effects from energy programs. The following table presents the effects that are currently found in the literature that have some level of documentation of the effect as a result of an energy program. The authors make no assumptions about the accuracy or validity of these effects, but present them to suggest the potential range of effects. In reviewing Table 11.2, it should be noted that the reverse effects should not be eliminated as a potential program result when establishing evaluation methods, tools, or in conducting the analysis.

Table 11.2: Partial Listing of Non-Energy Effects

Potential Effect	Definition
Ability to pay utility bills and reductions in utility collection costs	Program-induced energy savings can lower energy bills and influence the ability of participants to pay utility bills, leading to a reduction of utility collection costs.
Changes in features/options of energy efficient measures	Program measures may have more or less features and options than the less efficient measures being replaced.
Economic stimulus impacts	Program spending may have a net local economic effect from the program's materials and labor. Participant savings may have an economic impact as participants spend their "savings" on other goods and services. Net job creation may lower unemployment costs. As trade allies participate in programs, they may increase their product sales resulting in job creation and retention. Energy efficiency improvements may allow manufacturing plants to retain or hire more workers due to increased productivity or profitability associated with the energy efficient technologies.
Fewer customer callbacks	Trade allies that participate in programs may experience fewer customer callbacks related to problems with less efficient equipment or building features that may not be as well-designed, constructed or installed. Alternatively, new energy efficient technologies may have increased callback rates compared to standard equipment.
Greater market share for vendors and increased sales	Vendors who offer energy efficient products and services can capture greater market share, and increase sales if they are able to take business away from competitors that do not offer energy efficient products and services.

Table 11.2: Continued

Potential Effect	Definition
Improved comfort - draftiness, thermostat setting	Program measures, such as insulation and weather-stripping, may reduce drafts or the need to increase thermostat settings, thereby improving comfort.
Improved energy and energy-related perceptions, recognition, choices, or behaviors	A program may provide educational components such as written materials and on-site demonstrations. These features may improve participant knowledge and allow them to make better energy efficiency decisions. Programs may improve energy-related perceptions, ENERGY STAR® brand recognition, and their energy choices and behaviors.
Improved home appearance/aesthetics	Program measures, such as energy efficient appliances and equipment may improve the condition or appearance of the home, adding value.
Improved learning	Students may learn more when in an environment that is more energy efficient and more aesthetically productive to the learning environment.
Increased customer loyalty	Vendors and utilities that offer energy efficient products and services may be able to retain customers more readily.
Increased customer satisfaction	Vendors and utilities that offer energy efficient products and services may experience increased levels of satisfaction from their customers.
Increased product or service quality	Energy efficiency improvements are often associated with improved product or service quality.
Increased productivity	Energy efficient technologies and practices may also increase productivity.
Increased profitability	Trade allies that participate in programs may experience increased profitability.
Increased worker satisfaction and morale	Energy efficient technologies and practices can result in greater worker satisfaction.
Lower job-related injuries or illness	Energy efficient technologies and practices may result in fewer workplace injuries and illnesses.
Noise reduction - inside and from outside the home	Program measures such as insulation and weather-stripping may reduce noise being transmitted from the outside to inside the home.
Public relations or image	Energy efficiency may improve corporate image leading to customer attainment or attraction.
Reduced air emissions	Energy efficiency can reduce emissions and pollutants, contributing to global warming and acid rain, or to human health improvements.

Table 11.2: Continued

Potential Effect	Definition
Reduced indoor poisonous gases such as Carbon Monoxide (CO) and Carbon Dioxide (CO ₂)	When CO ₂ reduction measures are included in a program, health and safety effects can be accrued to the residents in terms of lower hospitalization costs and health-related expenses or avoided illnesses.
Reduced product losses	Energy efficient technologies may reduce product spoilage or result in fewer manufacturing defects or errors.
Reduced use of materials and/or waste and/or increased recycling	Energy efficient technologies and practices may reduce product and material waste, land filling and hazardous material handling costs.
Self sufficiency and improved household economics	Program-induced energy savings lowers energy bills and increases the economic security of participating households. This may lead to reductions in debt.
Water and waste water savings	Low-flow showerheads save energy and water, reducing water resource needs and aquatic impacts. Similarly, programs reducing water consumption via energy efficiency measures reduce the need for water treatment and sewage treatment facilities.

Making Decisions and Selecting Methods

This section presents the evaluation decision roadmap for determining if a non-energy effect evaluation is needed or can be funded for a specific program.

Steps in the Non-Energy Effects Evaluation Roadmap

Figure 11. describes the non-energy effects evaluation roadmap and helps to guide readers through the steps involved in making decisions about when NEEs evaluations can be conducted for California energy programs using Public Goods Charge funds. For programs operated through investor-owned utilities, the utility company administrating the programs typically makes these decisions. For programs operated through independent contractors, such as third-party programs, the lead program administrator typically makes these decisions.

Because California Public Goods Charge dollars are typically not spent evaluating NEEs, it is recommended that prior to conducting a NEEs evaluation the program administrator obtain permission from the CPUC to allocate program dollars to this research. Once permission is granted and an evaluation plan developed, the CPUC can approve or request modifications to the evaluation plan as appropriate. In budgeting the NEEs evaluation, the administrators will need to consider the sampling requirements and the threats to validity discussed in the Framework's Sampling and Uncertainty chapters (13 and 12, respectively).

In using the NEEs evaluation roadmap, decision makers should move through all of the decision steps in the roadmap to determine if a NEEs evaluation should be conducted. As a general rule, the NEEs roadmap suggests that programs needing NEEs documentation to increase program participation and related energy savings can conduct NEEs evaluations. However, these studies are not required by the CPUC. If the CPUC identifies one or more NEEs that need to be evaluated in order to document program effects, the CPUC can require specific non-energy program evaluations.

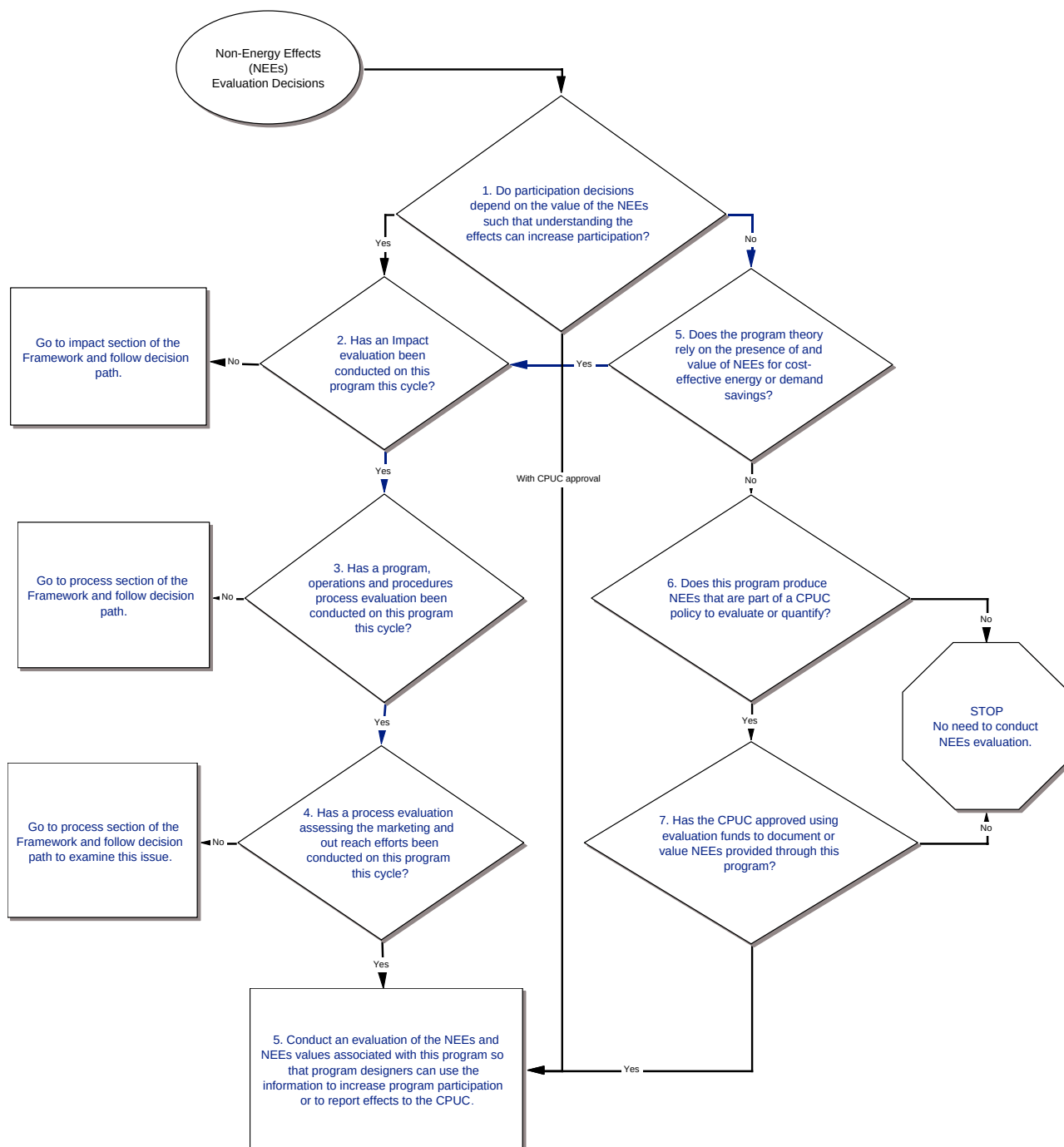


Figure 11.1: Non-Energy Effects Evaluation Framework Decision Path

The steps in the NEEs evaluation roadmap (Figure 11.1) are described below.

Step One

The first step is to determine if an evaluation documenting the presence and value of a non-energy effect can be used to help the program gain participants to increase the program's energy effects. Because the State of California, through the CPUC, is currently responsible for administering the Public Goods Charge funded programs and approval for the procurement programs, there is an inherent need that all programs provide accurate, reliable information in their outreach and marketing efforts. In addition, the CPUC realizes that there may be programs in which participation decisions rest more with the value of the non-energy effect than with the value of the energy impact associated with the program. While the primary goal of these programs is to achieve energy resources (directly or indirectly), it is realized that promotional information may need to be used that focuses on program benefits other than the energy savings. For this reason, programs that rely on the value of the NEEs to achieve participation can use evaluation resources to fund studies on the NEEs needed to increase participation.

The decisions associated with step one are either “no” (the program does not rely on the value of the NEEs to gain participation in order to accomplish its energy saving goals); or “yes” (the program does rely on the value of the NEEs). For programs that rely on the value of the NEEs, the program theory should clearly identify the importance of the NEEs in gaining energy impacts. In cases where the NEEs are used to accomplish energy savings goals, assessing the program theory and designing the evaluation accordingly become more important to be able to understand and document the linkages between the program intervention, the NEEs, the resulting behavior/actions obtained, and the resulting energy savings. (See the Framework section on program theory and logic models in Chapter 4 for more information and references.)

If the decision from this step is “yes,” then a NEEs evaluation can be a worthwhile addition to the evaluation effort. As a general rule the results from the step 1 decision will be yes or no. The paths from these two decisions are presented in the roadmap, and extend off the left or right of the roadmap's step 1 decision box. A “yes” to a step 1 decision moves the decision maker into a consideration for an impact or a process evaluation. This direction is meant to reflect the often higher priority placed upon impact and process evaluation if there are evaluation budget limitations that require a choice be made between the types of evaluations that can be conducted in any program cycle. However, in many cases it may be desirable to conduct the non-energy effect evaluation prior to the impact or process evaluations. In cases where the administrator thinks the program can increase the program's energy impacts beyond their stated goals, or they will not be able to reach their energy impact goals because of a lack of NEEs information, the administrator can request the CPUC to approve the use of funds to conduct the research and not wait until the impact and process evaluations are completed. For this reason a third decision branch is added to the step 1 decision that allows the administrator to move directly to a NEEs evaluation with CPUC approval.

Step Two

If the decision from step two is “yes,” and the program has energy impact goals, then the administrator may still need to conduct an energy impact evaluation of the program before funds are spent conducting a NEEs evaluation. If the program does not have specific energy impact goals, then an energy impact evaluation is not needed. This step is placed in the roadmap to make sure that the impact evaluation is placed in the highest priority position for the use of evaluation resources. If the program has energy impact goals and is in need of an impact evaluation (as a result of the Chapter 5 Umbrella Roadmap and current CPUC guidelines), then the impact evaluation takes precedent for the evaluation dollars. If an impact evaluation is not needed or has been conducted consistent with the impact evaluation roadmap (see Chapter 6: Impact Evaluation), then the process moves on to step three.

Step Three

This step is similar to step two except that the step three decision is associated with a process evaluation rather than an impact evaluation. The CPUC’s second highest priority for spending evaluation resources is for process evaluations. For this reason, the process evaluation is suggested for most programs. Step three asks if the program has conducted a process evaluation according to the requirements presented in the Chapter 5 Umbrella Roadmap and current CPUC guidelines. In this case, the process evaluation identified in step three is one that examines the program’s design, management, operation and implementation procedures to help the program improve its energy impacts or to become more cost-effective. If a process evaluation has not been conducted for this program, then a process evaluation is suggested prior to a NEEs evaluation. If a process evaluation has been completed in accordance with the process decision roadmap (see Chapter 8), then a NEEs evaluation may be conducted. (It should be noted here that the process evaluation roadmap suggests a number of conditions in which an administrator may not need or want to conduct this type of a process evaluation.)

Step Four

This step is similar to step three, but focuses on a decision directly associated with a specific kind of process evaluation. The step four decision focuses on whether a process evaluation has been conducted on the program’s marketing and outreach efforts. For most programs, the success of the program can be directly linked with the success of the program’s marketing efforts. If a program is not marketed well, then participation can suffer, causing difficulties in achieving the desired energy impacts. This step is placed in this roadmap to bring attention to the potential need for the process evaluation to examine and assess the primary program characteristics that have a direct bearing on the success of the program. The marketing and outreach process evaluation focuses on how the program is marketed, including the targets selected for the marketing approach, the materials used, and the results of these efforts. If the process evaluation conducted for the program included an assessment of the program’s marketing and outreach efforts, then a NEEs evaluation may be conducted. If the process evaluation did not assess the marketing and outreach efforts, then the program may want to conduct this type of process evaluation prior to conducting the NEEs evaluation. (It should be noted here that the process evaluation roadmap suggests that if a program is achieving its marketing and

enrollment related goals the program administrator may not need or want to conduct this type of a process evaluation.)

Step Five

This step is contingent on the decision from step one. If the step one decision is that “no,” the program does not depend on the value of the NEEs in order to increase participation and enrollment rates, then the decision process moves to step five. In this step the decision relates to whether the *program theory* relies on the presence and value of one or more NEEs in order to achieve cost-effective energy or demand savings. While step one asks about reliance on the non-energy effects for gaining participation in the program, the step five decision rests on whether the program theory indicates that cost-effective energy or demand savings rests on or is heavily influenced by the value of the NEEs. This decision presents a somewhat different approach to determining if the value of the non-energy effect is a critical or important step in the ability of the program to achieve energy savings, regardless of whether enrollment or participation in the program is influenced by the presence of the non-energy effect. While enrollment in a program may be consistent with the expectations of the program designers, this does not mean that all appropriate energy-saving technologies associated with that program will be adopted by participants. In some cases, enrollment may be high, but technology adoption for one or more measures may be lower than expected because of lack of information about the full benefits of the technology. If the program expects that a NEEs evaluation can produce information that could be used to increase adoption rates such that the program would be more cost-effective at obtaining its energy impact goals (excluding the value of the NEEs), then a NEEs evaluation may be appropriate. If the program theory indicates that it relies on the value of the non-energy benefits, then the decision process moves back to step two. If the step five decision is that “no,” the program theory does not rely on the presence of the non-energy effect to be successful, then the decision moves to step six.

Step Six

This step recognizes that the CPUC may identify NEEs that are in need of evaluation. This step allows that event to occur without influencing the energy impact evaluation or the program process evaluation efforts. If the CPUC has identified a specific non-energy effect that it wants evaluated, it will convey that need to the program administrators, contractors, or evaluators as appropriate.

Step Seven

This step is a “safety step” in the non-energy benefit process that requires the CPUC to specifically approve the use of evaluation funds for the purpose of researching CPUC identified NEEs. This allows the CPUC to have more control over the dollars spent for non-energy research to support a CPUC information need and direct control over the type of research to be conducted. If the CPUC has approved the use of evaluation funds to conduct the non-energy effect research, then a NEEs evaluation can be planned and conducted.

Conducting the Non-Energy Effects Evaluation

If any of the above detailed decision steps identify the need for a NEEs evaluation, the program can budget and plan for the implementation of the evaluation by an independent program evaluator that meets the skills requirements presented in this chapter and as is appropriate for the type of NEEs research to be conducted.

Chapter 12: Uncertainty

Introduction and Key Issues

This chapter discusses three topics that are especially relevant to evaluation contractors and policy makers alike:

- Bias and statistical precision,
- Techniques for assessing the savings in a portfolio of programs, and
- Allocation of evaluation resources to programs.

Periodically these topics have been discussed in energy efficiency evaluation but they are not a standard part of program evaluations or evaluation planning. Since evaluation is designed to reduce uncertainty, this chapter addresses several issues that are pivotal to successful evaluation. It applies to all types of evaluations: impact, process, market, education, etc. Consideration of areas of uncertainty, level of precision or confidence in the results, and potential bias occurs in all high quality evaluation plans and within the analysis and reporting of all evaluation results.

This chapter is especially important for policy makers. Evaluators and evaluation reviewers also need to be familiar with this material. The sections discussing “How Bias Arises” and “Reporting Guidelines” will give program managers and implementers an overview of some of the issues that arise in evaluation studies.

The chapter makes the case that the research design and data collection methodology are just as important to the quality and accuracy of an evaluation study as the size of the sample and the choice of analytical techniques. A sound evaluation study depends on all these elements but budget limitations may force tradeoffs. Unfortunately, it is generally difficult to quantify these tradeoffs, so effective evaluation requires a combination of science and experience.

Skills Required for the Uncertainty Analysis Tasks

This chapter is written from a statistical science perspective to understand relative precision and bias in a quantitative manner. Either making assessments or conducting quantitative analysis in this manner generally requires knowledge of basic graduate statistics or equivalent study and experience with a mentor in this area.

Mitigating the risk of bias requires an understanding of research or evaluation design. Relevant graduate courses or similar training by an experienced professional could be from within many of the social sciences (sociology, economics, psychology, policy

analysis, public health, etc.). Reducing potential measurement bias and response bias in surveys also requires training and experience in survey and instrument design.²⁷⁶

Bias and Statistical Precision

An evaluation study is designed to assess one or more parameters such as the gross and net annual savings of all projects undertaken in a program. A sound evaluation study should produce estimates that are free of any substantial bias and should accurately characterize the statistical precision of the results.

Most evaluators are familiar with the most commonly used techniques for reporting statistical precision: error bounds, confidence intervals, and relative precision. Some of us may not realize, however, that these techniques are very misleading if there is substantial bias in the findings. For example, it is misleading to report that the savings of a program have been measured within $\pm 10\%$ at the 90% level of confidence if there is a plausible risk that the results are biased by 25%.

Unfortunately, it is usually extremely difficult to objectively quantify the magnitude of the bias or even its direction. Occasionally, there is an opportunity to carry out a special methodological study designed to assess the risk of bias from one or more particular factors. But generally the best the users of an evaluation study can do is to look for certain aspects of the implementation of the study that could increase the risk of bias – such as

Bayesian Statistics

This chapter has been developed from a *classical statistics* paradigm. At present, classical statistics is the foundation for most evaluation work. *Classical statistics* builds on the frequency school of probability and statistical inference. *Bayesian statistics* presents an alternative paradigm to classical statistics that builds on subjective probability.

When samples are large and there is little danger of bias, classical confidence intervals are generally equivalent to Bayesian inference. The authors have deliberately steered clear of classical concepts, such as p-values, that are disputed under the Bayesian approach. Therefore this chapter's discussions should be of value from either a classical or Bayesian point of view.

Classical and Bayesian statistical methodology generally diverge when there is limited experimental information. For example, a Bayesian approach might be used to develop a subjective assessment of the amount of bias. This type of analysis would extend the methodology discussed in this chapter. The subjective foundation of the Bayesian methodology is its major potential disadvantage in evaluation applications. In conveying the findings of an evaluation study, it is important to clearly identify any subjective inputs used in the analysis.

²⁷⁶ There are many sources of information on survey design. The Survey Research Methods section of the American Statistical Association web site has links to university courses on the subject as well as an introductory set of pamphlets called "What is a Survey?" (American Statistical Association 1995). Another source of information is the *Research Methods Knowledge Base* (Trochim 2001). Other references are: *Survey Methodology* (Groves et al. 2004); and *Methods for Testing and Evaluating Survey Questionnaires* (Presser et al. 2004).

poor response rates in a survey, deliberate exclusions of sample projects, poor measurement techniques in a field study, instability of coefficients in a regression analysis, etc.

In a high quality evaluation, those implementing the study would strive to mitigate the risk of bias and to honestly report any circumstances about the study that might increase the likelihood of bias. Unfortunately, it usually takes extra time and money to reduce the risk of bias, and the usual measures of the statistical precision of the results may not be improved at all. For example, in order to reduce the risk of non-response bias in a telephone survey, a substantial investment may be needed in more extensive training for the surveyors, more call backs, and perhaps to offer a financial incentive to each respondent.

It may be tempting to accept a higher non-response rate and divert these resources to a larger sample size since this strategy will almost certainly give a narrower confidence interval. This strategy can seriously compromise the integrity of a study. To make appropriate judgments in planning and executing sound evaluation studies and in interpreting their results, evaluators, reviewers, and those using evaluation results need to understand what bias is, how it can arise, and how it can undermine an evaluation study.

Basic Definitions

Let μ denote a particular parameter of interest such as the true average savings per project in the population of projects to be evaluated and let $\hat{\mu}$ denote the estimate of μ to be produced by a particular evaluation study. Before the evaluation study has been completed, $\hat{\mu}$ can be regarded as a random variable. For any given value of μ , $\hat{\mu}$ will have a probability distribution that reflects sampling variability, measurement error, response error, variation due to randomness in the model assumed to generate the data, etc. This is called the sampling distribution of $\hat{\mu}$.

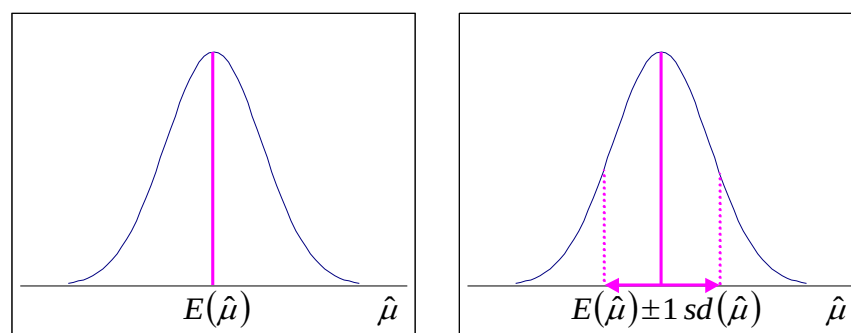


Figure 12.1: The Sampling Distribution of an Estimator

The two key parameters of the sampling distribution of $\hat{\mu}$ are the expected value of $\hat{\mu}$, denoted $E(\hat{\mu})$, and the standard deviation of $\hat{\mu}$, denoted $sd(\hat{\mu})$. Figure 12.1 shows an example of the sampling distribution of an estimator $\hat{\mu}$, together with the expected value

and standard deviation. In this example, the sampling distribution has been assumed to follow the normal probability distribution. The left side of Figure 12.1 shows the expected value of $\hat{\mu}$, $E(\hat{\mu})$. The right side of Figure 12.1 shows the one standard deviation interval around the expected value. If the sampling distribution of $\hat{\mu}$ follows the normal probability distribution,²⁷⁷ then the probability is about 68% that $\hat{\mu}$ will be in the interval $E(\hat{\mu}) \pm 1 \text{ } sd(\hat{\mu})$, i.e. in the interval denoted by the two-headed arrow shown in the right hand side of Figure 12.1.

Assuming a normal probability distribution, the probability is about 90% that $\hat{\mu}$ will be within 1.645 standard deviations of the expected value of $\hat{\mu}$, i.e., in the interval $E(\hat{\mu}) \pm 1.645 \text{ } sd(\hat{\mu})$. This is the usual basis for a 90% confidence interval. In particular, the quantity $1.645 \text{ } sd(\hat{\mu})$ is called the expected error bound at the 90% level of confidence. Assuming that μ is not equal to zero, the expected relative precision at the 90% level of confidence is defined to be the expected error bound divided by μ : i.e.,

$$rp = 1.645 \frac{sd(\hat{\mu})}{\mu}$$

The estimator $\hat{\mu}$ is said to be an unbiased estimator of the population parameter μ if and only if the expected value of $\hat{\mu}$ is equal to μ , i.e., if and only if $E(\hat{\mu}) = \mu$. The left side of Figure 12.2 illustrates an unbiased estimator. In this case, although any particular realization of $\hat{\mu}$ will probably be either larger or smaller than μ , in repeated sampling the average value of $\hat{\mu}$ will be equal to μ .

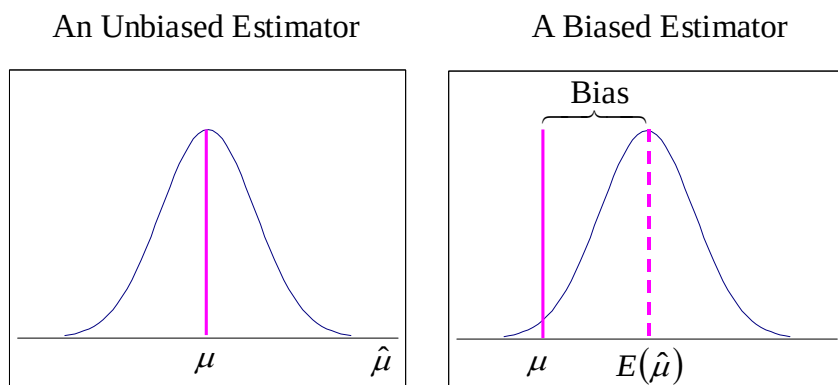


Figure 12.2: Unbiased and Biased Estimators

²⁷⁷ This will generally be approximately true for a well-designed evaluation study. The Central Limit Theorem and its various extensions suggest that if the sample is large the sampling distribution should generally be approximately normal. Even if the sample is moderate or small, this is generally true if an efficiently stratified sample design has been used. If a regression analysis is carried out with a small sample, it may be necessary to use special techniques, such as a Box Cox transformation, to improve normality. All of the material in this section can be found in most standard upper level undergraduate and graduate statistics textbooks.

If $\hat{\mu}$ is a biased estimator, the bias of $\hat{\mu}$ is the difference between $E(\hat{\mu})$ and μ . The right side of Figure 12.2 illustrates a biased estimator. In this example, $\hat{\mu}$ is so badly biased that the probability is high that $\hat{\mu}$ will be greater than μ . In other words, the evaluation study is very likely to overstate the actual value of μ .

A small amount of bias is usually not of concern, but the bias is serious if it is large compared to the standard deviation $sd(\hat{\mu})$. Assuming that μ is greater than zero, the relative bias is defined to be:

$$rb(\hat{\mu}) = \frac{E(\hat{\mu}) - \mu}{\mu}$$

The bias is serious if the relative bias is large compared to the expected relative precision. In this case, the relative precision and the error bound are likely to be misleading because they fail to reflect the bias.

When there is serious bias, a theoretically more suitable measure of statistical precision is the root mean square error of $\hat{\mu}$, given by

$$rmse(\hat{\mu}) = \sqrt{(E(\hat{\mu}) - \mu)^2 + sd(\hat{\mu})^2}$$

The root mean square error combines both the bias and the standard deviation, and is often used in specialized studies, especially simulation studies. If the root mean square is available, then the generalized error bound can be defined to be $1.645 rmse(\hat{\mu})$. If in addition $\mu > 0$ then the generalized relative precision can be defined to be $1.645 rmse(\hat{\mu})/\mu$. In most evaluation studies, however, it is difficult to quantify the bias so the root mean square error is not reported. Generally, the reported error bound and relative precision ignore any possible bias.

Table 12.1: Example of a Biased Estimator

True Value	1000	
Expected Value	1250	
Standard Deviation	100	
Bias	250	1250 - 1000
Relative Bias	0.25	250 / 1000
Root Mean Square Error	269	$\sqrt{100^2 + 250^2}$
Generalize Error Bound	443	(1.645) (269)
Generalized Relative Precision	0.44	(1.645) (269) / 1000
Reported Error Bound	165	(1.645) (100)
Reported Relative Precision	0.16	165 / 1000

The example given in Table 12.1 illustrates some of these ideas. In the example it is assumed that the true parameter of interest is 1,000 units. The estimator produced by the

study is assumed to have a normally distributed sampling distribution with an expected value of 1,250 units and a standard deviation of 100 units. In this case, the probability is about 90% that the estimator will be in the interval $1,250 \pm 165$, i.e., from 1,085 to 1,415, although the true value is 1,000. So the estimator is likely to be too high.

This estimator is expected to perform so badly because it has a bias of 250 units. In this case the root mean square error is 269 units, which would give a generalized error bound of 443 units and a generalized relative precision of 44%. If, as is the common practice, the bias were ignored in planning the study or in reporting the statistical precision, the expected error bound would be 165 units and the expected relative precision would be 16%. In short, if bias is ignored, the estimator would be expected to be within 165 units of the true parameter, whereas in actuality, the estimator is likely to be between 85 and 415 units higher than the true parameter.

How Bias Arises

Most evaluation studies are free of substantial bias, as long as the underlying assumptions are accurate. But bias can arise when the assumptions break down in the implementation of the study. This breakdown of assumptions can be spotted, but only by thoroughly understanding the underlying assumptions and being close to the details of the study. Thus the principal investigators are generally in the best position to discuss any risks of bias.

Assessing all of the various risks is a challenge, even to the most experienced study director. The violation of almost any underlying assumption can lead to bias. So there are as many potential sources of bias as there are assumptions in a study.

Some of the most common sources of potential bias in energy efficiency evaluation include the following:

- **Non-response and other forms of selection bias.** As discussed in the Sampling chapter, a key assumption in statistical sampling is that the sample is randomly selected from a complete and accurate sampling frame following a suitable sampling plan. Selection bias can arise if the sampling frame or sampling plan excludes a particular part of the target population. Non-response bias can occur if any of the designated sample projects are dropped or replaced for almost any reason, such as refusal to participate, technical difficulties, etc.^{278, 279} In some engineering field studies, the sample is sometimes subjectively selected. This practice not only raises the risk of bias but also makes it nearly impossible to assess the statistical precision of the estimator.

²⁷⁸ Occasionally, the target population is a subset of the sampling frame that can only be identified from the study itself. This should not be confused with non-response.

²⁷⁹ All estimators commonly used in survey sampling will generally be essentially unbiased as long as the case weight applied to each sample project is equal to the reciprocal of the probability that the project is included in the sample, i.e., the true inclusion probability, considering all factors including the sample design, any non-response, technical exclusion, etc.

- **Measurement bias.** Any consistent difference between the value of a measurement and the underlying property can lead to measurement bias. There are as many causes of measurement bias as there are measurement techniques, e.g., poorly worded survey questions, inability or reluctance of the respondent to provide the desired information, inappropriate scoring of survey responses, oversimplified engineering models, inaccurate calibration of measurement instruments, etc.
- **Erroneous specification of the statistical model.** An erroneous statistical model can threaten a statistical evaluation study just as seriously as an erroneous engineering model or badly implemented sampling plan can threaten a field study. Many evaluation studies make heavy use of statistical regression analysis. Any violation of the assumptions of regression analysis can introduce the risk of seriously misleading findings.
- **Choosing an appropriate baseline.** In many evaluation studies, there is more uncertainty about the baseline energy consumption than about the post-participation energy consumption. In evaluating an AC rebate program, for example, it is necessary to decide whether the program participant would have retained the existing air conditioner in the absence of the program or replaced it anyway. If the latter, how efficient would the replacement unit have been in the absence of the program? Similarly, in evaluating a new construction program, is it appropriate to assume that the building would have been built to a particular energy code?
- **Self selection of program participants.** Most common methods of assessing free ridership compare the behavior of program participants to a sample of non-participants. The non-participants are generally assumed to reveal what the behavior of the participants would have been in the absence of the program. These methods can provide biased estimates of free ridership if the underlying assumptions are not accurate, e.g., if there is a correlation between interest in conservation and program participation.
- **Misinterpretation of association as causal effects.** Most introductory texts on regression and econometrics discuss the danger of interpreting a regression coefficient as an unbiased estimator of the causal effect of the explanatory variable on the dependent variable. This type of bias is always a danger when regression and other statistical techniques are used to analyze data that has not been produced in a controlled experiment. These concerns are relevant in those impact evaluation studies that seek to estimate the true impact of a program on energy consumption by analyzing the tracking and billing data collected in the program itself.

The research design literature sometimes discusses various threats to the validity of a study using terms such as the following.²⁸⁰

²⁸⁰ Two examples of this literature are: *Applied Social Research* (Rubin 1983) and *Evaluation* (* Weiss 1998).

- **Construct validity.** The degree to which the inferences are valid based upon the way in which the underlying construct being tested is operationalized in the research design.²⁸¹
- **Statistical validity.** Using an incorrect statistical methodology in the analysis, e.g., using ordinary least squared regression analysis when there is substantial random error in an explanatory variable.²⁸²
- **Internal validity.** Inadequate controls in the research setting (or quasi-experimental design), instrumentation and selection biases.
- **External validity.** Can the evaluation findings be applied to the entire treatment group?²⁸³

These issues are closely related to the sources of bias that have been discussed such as non-response, measurement error, error in specifying the statistical model, etc.

Special Techniques to Assess Bias

Since bias can be associated with the breakdown of almost any underlying assumption, it is difficult in a typical evaluation study to quantify all potential sources of bias or even to discuss all the different ways that bias might arise. The analysis itself is guided by the underlying assumptions and it is usually difficult to carry out the analysis that would be appropriate in the absence of these assumptions or under alternatives to them. Moreover, alternative assumptions usually lead to different conclusions.²⁸⁴ Usually the director of an evaluation study must decide which assumptions are most plausible and to emphasize the corresponding results.

However, when it is feasible, a high quality study would report the results obtained under all plausible alternative assumptions. This is generally more practical in a regression study than in a survey or field study. This often takes the form of reporting regression models with alternative choices of explanatory variables and perhaps alternative ways of dealing with outliers.

In surveys and other applications of sampling, it usually is difficult to assess how much bias might be caused by non-response since there usually is limited data for the non-respondents. If simple random sampling has been used, it sometimes is useful to compare the tracking estimates of savings of the respondents to the population since a

²⁸¹ See *The Research Methods Knowledge Base* (Trochim 2001).

²⁸² Bayesian statisticians would include the reliance on p-values. This methodology may lead to false conclusions that can appear quite differently in Bayesian analysis. For example, see the article: *Facts versus Factions: The Use and Abuse of Subjectivity in Scientific Research*. (Matthews 1998).

²⁸³ A classic graduate text dealing with research design is *Experimental and Quasi-Experimental Designs for Research* (Campbell and Stanley 1963). They present extensive discussion on various quasi-experimental designs and also discuss eight types of threats to internal validity and four to external validity.

²⁸⁴ One indicator of a serious problem is if plausible alternative assumptions yield confidence intervals will little or no overlap.

significant difference would indicate that the sample is biased. Another approach to investigating the effect of non-response bias is to carry out a special study of a sample of the non-respondents. In practice, however, it usually is difficult to collect meaningful information from non-respondents.

If stratified ratio estimation is used to estimate the program realization rate (the ratio of evaluation estimated savings to program tracking estimated savings, the degree to which estimated savings are “realized”), then it is generally difficult to estimate the effect of non-response bias. In theory, if an efficiently stratified sample design is used, if the case weights are the reciprocal of the true inclusion probabilities, and if there is negligible measurement bias, the stratified ratio estimator will be essentially unbiased. Bias can arise if the case weights fail to accurately reflect the inclusion probabilities due to non-response or other selection factors. In principle, one or more statistical models could be developed for the probability of response, the response probabilities could be factored into the case weights, and the sensitivity of the estimates to the changes in the case weights could be explored. But this work would divert resources from the evaluation study itself. Usually it seems better to devote these resources to minimizing non-response in the first place.

The Sampling chapter discusses the use of a ratio model to assist in the development of an efficient sample design. It is important to note that the stratified ratio estimator is still generally unbiased even if the assumed model is inaccurate – provided that the sample design has been followed. Moreover, the calculated statistical precision will generally be a good guide to the statistical precision of the stratified ratio estimator. In other words, the stratified ratio estimation approach is designed to provide better protection against model-specification bias than econometric methods – provided that the sample design is accurately followed.

However, these results assume large samples whereas in many evaluation studies stratified ratio estimation is used with small or moderate sample sizes. So these results may not be entirely applicable. A special simulation study could be carried out to assess whether stratified ratio estimation techniques are seriously biased when used with moderate sample sizes, especially when the ratio model is not accurate. In practice, however, it is often hard to draw general conclusions from this type of simulation study.²⁸⁵

Sometimes a special study can be carried out to assess measurement bias. If a particular aspect of a field study is challenged, an independent study can sometimes be implemented to validate the data collection and analysis. For example, short-term metering might be used to validate the accuracy of reported schedules of use in a small sample of projects. Often the study is carried out in a sub-sample of the original sample so that the new results can be compared to the information collected in the original study.

²⁸⁵ One large simulation study showed that stratified ratio estimation had little or no bias in several load research applications. See *Sample Designs for Load Research: The Bootstrap Comparison Procedure*. (McCarthy et al. 1985).

Ledyard reported an example in which end-use metering was used to verify the operating hours of lighting measures reported in an on-site survey.²⁸⁶

An excellent analysis of measurement error can be found within a billing analysis performed by Rick Ridge of Ridge & Associates for Pacific Gas & Electric Company. In a statistically-adjusted engineering (SAE) billing analysis, the estimated realization rate can be biased if the tracking system estimates of savings used in the analysis have random error and systematic error.²⁸⁷ This paper cited earlier work²⁸⁸ that demonstrated that measurement error in a single explanatory variable would cause the coefficient to be downwardly biased. With multiple explanatory variables, the coefficients can be biased in unknown directions. A Monte Carlo simulation was done utilizing specific systematic error of plus and minus 10% and several forms of random error.

This work verified that the SAE methodology worked well with systematic error; the coefficient adjusted for the systematic error as expected. The work showed, however, that with random error of 15% or greater, the SAE analysis performed worse than using a dummy variable in the billing analysis, evidently because the dummy variable is free of error. The importance of this was shown in the billing analysis of PG&E's 1994 Commercial HVAC program. The SAE realization rate was 0.35 using the tracking estimates of savings. When the tracking estimates were restricted to those expected to have less measurement error, the realization rate was found to be 0.50. However, with the dummy variable approach that is free of measurement error, the realization rate was 0.92. Given these results, Ridge recommended that using a dummy variable may be preferable to the SAE methodology for HVAC impact analyses unless the engineering priors are free of significant random error. The dummy variable coefficient may not be as precise but it is free of the bias associated with the SAE coefficient so that it appears to be more reliable.²⁸⁹

Another interesting assessment of the importance of bias versus precision can be seen in a simulation study on this topic conducted by Richard Sonnenblick and Joseph Eto from Lawrence Berkeley Laboratory in 1995. Their work found that "imprecision in the cost of conserved energy was significant for programs with mean TRC test ratios close to one, while higher ratios guarantee cost-effectiveness even with considerable estimate imprecision." They stated that "a 90/10 criteria for precision seems excessive for most programs." On the other hand, bias in savings estimates had a greater impact where bias could threaten cost-effectiveness for programs with ratios approaching 2.0. From this, they concluded that "much of the contemporary concern with precision should be redirected to examine bias in evaluation estimates."²⁹⁰

²⁸⁶ "Evaluating the Underserved Small C&I Market: Building a Bridge to Implementation." (Ledyard 2003).

²⁸⁷ SAE billing analysis is described in the Billing Analysis section of Chapter 6: Impact Evaluation.

²⁸⁸ *Econometrics: Basic and Applied* (Johnson et al. 1967).

²⁸⁹ "Errors in Variables: A Close Encounter of the Third Kind." (* Ridge 1997).

²⁹⁰ "Calculating the Uncertainty in Estimates of DSM Program Cost-effectiveness." (* Sonnenblick and Eto 1995). Quote on page 767.

Reporting Potential Bias

Under ideal conditions, all common evaluation approaches yield essentially unbiased estimators – but conditions are rarely ideal. Therefore it seems likely that most evaluation studies have some amount of bias. Unfortunately, it is extremely difficult to objectively quantify the possible bias. Given that bias is generally immeasurable; evaluation reports generally focus on precision, which is measurable.

The effect of bias, however, can be quite important if it exists. You could be quite precise about the answer, but it would be the wrong answer. Given this, potential bias is critically examined in any high quality evaluation planning and evaluation analysis. Greater evaluation resources spent for larger sample sizes to achieve greater precision will only provide more misleading results if the study is biased. In planning and implanting evaluation studies, addressing sources of bias is as important as increasing precision.

As indicated above, the principle investigators are generally in a better position than anyone else to discuss the risk of bias. Therefore, the evaluators are best able to help the reader assess the danger of bias. A sound evaluation would at a minimum discuss all relevant sources of bias and potential threats to validity – non-response, measurement, model specification, self-selection, etc. When appropriate, the report would also discuss the results obtained under plausible alternative models.²⁹¹

Reducing the Standard Deviation of an Estimator

The standard deviation of an estimator depends on all aspects of a study including the general experimental plan, the sampling plan, the sample size, the data collection and verification techniques, and the statistical analysis methodology. In planning an evaluation, it is necessary to consider all of these factors.

For example, in the chapter on sampling it is noted that the standard deviation of a stratified ratio estimator is often much smaller than the standard deviation of the sample mean of a simple random sample of observations. The improvement in statistical precision comes partly from changing the focus from the average actual savings per project to the realization rate, and partly from taking advantage of the supporting information provided by the tracking system to develop an effective sample design. Stratified ratio estimation can often be carried out at little added cost compared to simple random sampling, but with a substantial improvement in statistical precision or reduction in the sample size. The reduction in the sample size, in turn, can allow greater attention to mitigating bias.

Similarly, if a regression analysis of billing data is to be used to evaluate a retrofit program, it is necessary to consider all aspects of the study, e.g., how many months of

²⁹¹ A good discussion of reporting requirements is provided in *Quality Assurance Guidelines for Statistical, Engineering, and Self-Report Methods for Estimating DSM Program Impacts*. (* CADMAC 1998).

billing data should be required before and after the retrofit, how closely the non-participants should be matched to the participants, how much the billing data should be cleaned prior to analysis, how should weather data be used in the analysis, whether a survey should be carried out to provide information about other factors affecting energy consumption during the study period, etc. All of these issues are likely to affect the standard deviation of the resulting estimate of savings. In particular, to the extent that variation can be controlled without increasing the risk of bias, the statistical precision can often be substantially improved from an effective experimental approach, often with little added cost.²⁹²

Integrating the Results from Multiple Evaluation Studies

One of the paramount objectives of evaluation is to provide an accurate assessment of the total savings achieved by a set of programs. For example, policy makers may want to talk about the actual savings and cost-effectiveness of:

- All conservation programs taken together compared to supply-side options,
- Residential programs compared to non-residential programs,
- New construction programs compared to retrofit programs, and
- Third-party programs compared to utility-delivered programs.

To make these comparisons it is necessary to assess the aggregate savings and the corresponding statistical precision of various portfolios of programs and to assess whether there is a significant difference in the savings of programs.²⁹³ To do this, it is usually necessary to combine the results of two or more independent evaluation studies. In addition, policy makers may have occasion to contrast and combine the results of independent studies of the same program. The following techniques can be used individually or in combination to address most situations.²⁹⁴

- Estimating the Total Savings of a Portfolio of Programs,
- Contrasting the Results of Two Independent Studies,
- Pooling Two Statistically Independent Estimators of the Same Parameter, and
- Chaining the Results of Two Evaluation Studies.

²⁹² Special techniques may be needed to correctly estimate the standard deviation of a complex estimator when there are several sources of variation such as sampling, modeling, etc. Chapter 11 of *Model Assisted Survey Sampling* (Sarndal et al. 1992) discusses several methods that are sometimes useful in evaluation, including balanced half sample methods, the Jackknife technique, and Bootstrap analysis. There is a rich literature on methods such as these.

²⁹³ However, any such comparison should take into consideration that some energy efficiency programs may be handicapped by participation requirements. These and other factors may need to be taken into account to ensure an appropriate comparison, i.e., an “apples to apples” comparison.

²⁹⁴ These and other methods were discussed in the context of evaluation in “A Methodology For Integration of Evaluation Studies” (Wright and Jacobson 1993).

Each of these techniques will be discussed in the following sections. Throughout this discussion, it is assumed that each study has provided an unbiased estimate of the parameter of interest and an accurate error bound or relative precision. A later section in this chapter will also show how to calculate the error bound or relative precision when a p-value is given.

Estimating the Total Savings of a Portfolio of Programs

Suppose the total savings from a set of programs is desired. Consider the following assumptions:

1. There are no interactions between the programs,
2. Each of the individual programs has been evaluated independently,
3. Each evaluation has provided an unbiased estimator of the actual savings of the corresponding program, and
4. Each evaluation has provided the corresponding error bound.

Then the total savings of the portfolio can be estimated as the sum of the savings of each program. This is an unbiased estimator of the total savings. The corresponding error bound is the square root of the sum of the squares of the error bounds of each program.

Table 12.2 provides an example. In this hypothetical example, a portfolio of four programs is being examined. Programs A and B are relatively small, and each have estimated savings of 50,000 units. Program C is twice as large as A and B, with an estimated savings of 100,000 units. Program D is five times larger than program C, with an estimated savings of 500,000 units. It is assumed that there are no interactions between these programs so the total savings of the portfolio can be estimated as the sum of the estimated savings of the four individual programs, i.e., 700,000 units.^{295, 296} The evaluation contractors also need to have accurately reported the error bound of the estimated savings of each of the four programs at the 90% level of confidence, as 1.645 times the actual standard deviation of the estimated savings. These results are shown in the last column of Table 12.2. Another important assumption is that the four estimates are mutually independent random variables.²⁹⁷ The corresponding relative precision is shown for completeness.

²⁹⁵ Two programs could interact, for example, if they both provided the same measures to the same participants. A more subtle form of interaction could occur if one program affected the free ridership of another program by making a technology more familiar in the market.

²⁹⁶ The method of pooling generally can be used for both gross and net savings.

²⁹⁷ The result is a simple consequence of (a) the fact that the standard deviation of a sum of statistically independent random variables (e.g., the estimated savings of each program) is the square root of the sum of the squares of the standard deviations of each of the random variables, and (b) the error bound being defined as 1.645 times the standard deviation.

Table 12.2: Pooling the Savings of a Portfolio of Programs

Program	Estimated Savings	Relative Precision	Error Bound
A	50,000	13%	6,264
B	50,000	10%	4,852
C	100,000	6%	6,264
D	500,000	4%	18,792
Total	700,000	3%	21,334

In this example, the error bound associated with the total estimated savings is easily calculated as the square root of the sum of the squared error bounds of each of the four programs: $\sqrt{6,264^2 + 4,852^2 + 6,264^2 + 18,792^2} = 21,334$. The 90% confidence interval for the savings of the portfolio is $700,000 \pm 21,334$ units. The relative precision of the total portfolio savings is $21,334/700,000 = 3\%$.

Note that in this example, the relative precision of the total savings is better than the relative precision of any of the four programs. This is a direct consequence of our assumption that the four estimators are mutually independent and free of bias.

Contrasting the Results of Two Independent Studies

A common question is whether there is a significant difference between two programs. For example, a policy maker may wish to compare the realization rate achieved by two different programs. Assume that there are statistically independent and unbiased estimates of the realization rates of each of the two programs, and that the associated error bounds have been provided. Then the error bound associated with the difference between the two realization rates can be calculated as the square root of the sum of the squared error bounds of each of the two programs. If the observed difference is greater than the error bound associated with the difference, then the difference is statistically significant at the 90% level of confidence. Otherwise the observed difference is not statistically significant.²⁹⁸

Table 12.3 shows an example. In this example, independent evaluation studies have been carried out for two programs, A and B. Program A was found to have a realization rate of 0.80 whereas program B was found to have a realization rate of 0.70. Each of these results is assumed to be unbiased, and the error bounds are 0.12 and 0.10 respectively. Given these results, do the two programs have a *significantly* different realization rate? In other words, is the observed difference statistically significant?

²⁹⁸ This result is a simple consequence of (a) the fact that the standard deviation of the difference between two statistically independent random variables (e.g., the estimated savings of each program) is the square root of the sum of the squares of the standard deviations of each of the random variables, and (b) the error bound is 1.645 times the standard deviation.

To address this question, the error bound associated with the difference between the two realization rates is calculated as the square root of the sum of the squared error bounds of each of the two programs: $\sqrt{0.12^2 + 0.10^2} = 0.16$. This implies that the 90% confidence interval for the difference in the two realization rates would be 0.10 ± 0.16 or from -0.06 to +0.26. Since the error bound is greater than the difference itself, it should not be concluded that the difference is statistically significant in this example.

Table 12.3: Contrasting Program A and B

<i>Program</i>	<i>Realization Rate</i>	<i>Relative Precision</i>	<i>Error Bound</i>
A	0.80	15%	0.12
B	0.70	14%	0.10
Difference	0.10		0.16

Table 12.4 shows a second example. In this case the realization rate of Program A is being compared to that of Program C. As the table shows, the difference in the two realization rates is 0.25 and the error bound is 0.18. In this case the difference is greater than the error bound and so the difference is statistically significant.

Table 12.4: Contrasting Program A and C

<i>Program</i>	<i>Realization Rate</i>	<i>Relative Precision</i>	<i>Error Bound</i>
A	0.80	15%	0.12
C	0.55	25%	0.14
Difference	0.25		0.18

The preceding methods can also be used to contrast the results of two statistically independent studies of the same program. This type of application is discussed next.

Pooling Two Statistically Independent Estimators of the Same Parameter

Sometimes, two statistically independent evaluation studies are carried out for the same program. For example, the first study might use billing analysis techniques and the second study might use on-site audits and engineering analysis of a sample of projects selected from the program. The first step is to check to see if there is a statistically significant difference between the results of the two studies, using the method previously discussed.

If the difference is not statistically significant, it may be plausible to assume that both studies have provided statistically unbiased estimates of the savings of the program.²⁹⁹ In this case, the results may be combined into a single estimate.

²⁹⁹ If there is a statistically significant difference between the results, it is implausible that both estimators are unbiased. In this case it is more reasonable to assume that one or both of the studies is biased. In this case, a substantial effort may be required to understand the source of the bias.

Assume that both studies have provided an unbiased estimator of a given parameter, such as the average savings per project, and that the error bound is known for each of the two estimates. What is the best way to combine the two estimates into a single estimate, and what is the associated error bound? The answer lies in the following rule:

If two estimators are both unbiased estimators of a given parameter, then any weighted average of the two estimators is also an unbiased estimator. To provide the minimum error bound, each of the two estimators should be weighted in proportion to the reciprocal of the squared error bound of the estimator. The error bound of the result is the square root of the reciprocal of the sum of the weights.

It may be helpful to express this rule algebraically. Let $\hat{\mu}_1$ and $\hat{\mu}_2$ be two independent, unbiased estimators of the average savings per project μ of a given program. Let eb_1 be the error bound of $\hat{\mu}_1$ and let eb_2 be the error bound of $\hat{\mu}_2$. The weights are defined as: $w_1 = 1/(eb_1)^2$ and $w_2 = 1/(eb_2)^2$. Then the pooled estimator of μ is calculated as the weighted average

$$\frac{w_1 \hat{\mu}_1 + w_2 \hat{\mu}_2}{w_1 + w_2}.$$

The corresponding error bound is $\sqrt{\frac{1}{w_1 + w_2}}$.

Table 12.5 and Table 12.6 give a numerical example. In this example two independent studies have yielded estimates of the average savings per project of a given program. The first study indicated that the savings were 600 units with an error bound of 100, and the second study indicated that the savings were 500 units with an error bound of 75. Before pooling the two results, it is generally wise to examine the difference between the two estimates, as shown in Table 12.5. The difference between the two estimates is 100 units with an error bound of $\sqrt{100^2 + 75^2} = 125$, so the difference is not statistically significant.

Table 12.5: Contrasting Two Estimators

<i>Estimator</i>	<i>Average Saving</i>	<i>Relative Precision</i>	<i>Error Bound</i>
1	600	17%	100
2	500	15%	75
Difference	100		125

Therefore it is plausible to assume that the two results are both unbiased estimators of the true average savings of the program. Under this assumption, it is useful to pool the two estimates, as shown in Table 12.6. To carry out this calculation, the weight associated

with each of the two estimators is calculated as $w_1 = 1/100^2 = 0.0001$ and $w_2 = 1/75^2 = 0.0001778$. Notice that the second weight is larger than the first since the second estimator has a smaller error bound.

Then the pooled estimate is calculated using:

$$\frac{w_1 600 + w_2 500}{w_1 + w_2} = 536 \text{ units.}$$

Finally the associated error bound is calculated as $\sqrt{\frac{1}{w_1 + w_2}} = 60$ units. Since the pooled estimate is combining the information from the two studies, the error bound of the pooled estimate is smaller than the error bound of either of the two individual estimators.

Table 12.6: Pooling Two Estimators

<i>Estimator</i>	<i>Average Saving</i>	<i>Relative Precision</i>	<i>Error Bound</i>	<i>Weight</i>
1	600	17%	100	0.0001000
2	500	15%	75	0.0001778
Pooled	536	11%	60	0.0002778

Chaining the Results of Two Evaluation Studies

The method of chaining is relevant if the estimators of two factors are independent and unbiased, such as the gross realization rate and the net-to-gross ratio, and the product of the two factors is an estimate of interest. The relative precision of each of the two estimators is used to obtain the relative precision of the chained result. The relative precision associated with the product of the two factors is approximately equal to the square root of the sum of the squared relative precisions of each of the two factors.³⁰⁰

Table 12.7 shows an example. Assume that the gross savings of a program has been estimated to be 100,000 units with a corresponding relative precision of 15% at the 90% level of confidence. The net-to-gross ratio (NTGR) has been estimated in an independent study to be 0.8 with a relative precision of 15%. Then the net savings of the program is estimated to be 80,000 units. The corresponding relative precision is

$\sqrt{0.10^2 + 0.15^2} = 0.18$. The corresponding error bound is $80,000 \times 0.18 = 14,422$ units.

³⁰⁰ This equation can be derived by using Taylor's theorem to find a linear approximation to the product of the two factors near the expected values of each of the two factors, as explained further in our discussion on the propagation of uncertainty, below. (Most graduate and advanced statistics or econometrics textbooks use and refer to Taylor's theorem as a method to derive a linear approximation of a nonlinear function. For example see *Elements of Econometrics* (Kmenta 1971), pages 399-400.

Table 12.7: Chaining

<i>Factor</i>	<i>Estimate</i>	<i>Relative Precision</i>	<i>Error Bound</i>
Gross Savings	100,000	10%	10,000
NTGR	0.8	15%	0.12
Net Savings	80,000	18%	14,422

Converting a P-Value to a Relative Precision or Error Bound

When regression analysis is used to estimate the savings of a program, a so-called p-value is sometimes used to assess the statistical precision of the estimated savings. Most standard regression software packages report the estimate itself, the standard error, a statistic called the t-value, and the p-value. If the p-value is less than 0.10, the corresponding estimate is usually regarded as statistically significant at the 90% level of confidence.

To ensure that the results can be easily compared to the evaluation results for other programs, the statistical precision should also be reported as an error bound and relative precision at the 90% level of confidence. The error bound can usually be calculated as 1.645 times the reported standard error. The relative precision can be calculated as the error bound divided by the absolute value of the estimate.³⁰¹

Table 12.8: Converting a P-Value

<i>Variable</i>	<i>DF</i>	<i>Parameter Estimate</i>	<i>Standard Error</i>	<i>t Value</i>	<i>Pr > t </i>	<i>Error Bound</i>	<i>Relative Precision</i>
Intercept	1	493	163	3.03	0.0028	268	0.54
Pre-use	1	0.89	0.02	48.22	<.0001	0.03	0.03
Business	1	4,321	984	4.39	<.0001	1,618	0.37
Program	1	-536	199	-2.69	0.0078	328	0.61

Table 12.8 shows an example.³⁰² In this example, a multivariate linear regression model has been used to relate the post-retrofit annual energy consumption of 200 customers to their pre-retrofit annual energy use (pre-use), the change in business conditions for the customer (business), and participation in the program (coded 1 if yes, 0 if no).³⁰³ The column labeled Parameter Estimate shows the estimated intercept and the regression coefficients associated with each explanatory variable. The remaining columns show the corresponding standard error, t-value and p-value.

³⁰¹ Sometimes, the estimate of the impact of the program is based on more than one coefficient of the regression model. In this case special techniques may be needed to determine the appropriate standard error.

³⁰² Although the data used to create this example was simulated, the authors believe that this example is representative of many actual applications of regression analysis in impact evaluation.

³⁰³ In this example, one half of the 200 sample customers are program participants. The regression model explains 92% of the variance, i.e., $R^2 = 0.92$.

The regression coefficient associated with the program indicator variable is the primary variable of interest. If the usual assumptions of regression analysis are satisfied,³⁰⁴ this coefficient should be an unbiased estimate of the average reduction in annual energy consumption for each program participant. In this case, the regression results suggest that the program has a savings of about 536 units per participant. Moreover, since the p-value associated with this coefficient is very small (0.0078), an evaluator might conclude that this result is highly significant.

However, the error bound and relative precision suggest a different conclusion. Table 12.8 has two additional columns to show the error bound and relative precision of each parameter estimate. In the case of the program indicator variable, the error bound is calculated as $1.645 \times 199 = 328$ and the relative precision as $328/536 = 0.61$. Therefore a 90% confidence interval for the annual savings is 536 ± 328 units per participant. The error bound and relative precision statistics indicate that the statistical precision is rather poor, despite the excellent p-value of 0.0078. These results show that the p-value can give a misleading indication of the statistical precision of the results of a regression analysis.

Allocation of Resources to Evaluation

Policy makers must decide how much to spend on evaluating a portfolio of programs and how to allocate the spending to the individual programs in the portfolio. This is a complex issue and many factors can influence the analysis, including:

- The amount of savings expected from each program,
- Whether the program is expected to grow or shrink in the future,
- The uncertainty about the savings,
- How long it has been since the last evaluation and how much the program has changed in the interim,
- The unit cost of evaluating each sample project in the program,
- The variability of savings in the population of projects in the program, as measured by the coefficient of variation or the error ratio, as discussed in the chapter on sampling.

³⁰⁴ The authors have presented the linear regression model to simplify the exposition. The data were simulated assuming that the logarithm of post-use is linearly related to the logarithm of pre-use together with the business and program variables. Therefore our linear regression model is actually mis-specified. The conclusions are essentially the same if the properly specified regression model is used, with the logarithm of post-use as the dependent variable and the logarithm of pre-use together with the business and program variables as the explanatory variables. Under this second model, the standard assumptions are fully satisfied since the model accurately reflects the procedure used to simulate the data. The estimated coefficient of Program is -0.07. This indicates about a 7% reduction in energy consumption due to program participation. The p-value is 0.0396, which is highly significant. But the error bound is 0.06 and the relative precision is 0.79 so the statistical precision is not as good as might be suggested by the p-value.

The objective for this section is to discuss some statistical methodology that can help guide policy makers in these issues.

Bayesian Decision Theory and Power Analysis

Bayesian decision theory provides a structured methodology for determining the value of information and for making decisions between options for collecting information. For example, decision-tree analysis was used by Eric Hildebrandt of the Sacramento Municipal Utility District (SMUD) to assess the value of improved information from impact evaluation. Hypothetical scenarios were examined using cost and impact data for a typical commercial and industrial retrofit program. The decision-tree analysis looked at canceling or redesigning a program versus continuing a program given the information available and chance nodes for energy savings, demand impact, free riders, and the accuracy of the evaluation results. Including program costs and expected benefits allowed this analysis to assess the value of information. A scenario of a program with a break-even benefit-cost of 1.0 found the expected annual value of additional information to range from \$8,000 to \$182,000. The decision-tree framework was recommended “as a powerful and objective framework for setting impact evaluation budgets and priorities.”³⁰⁵

In principle, Bayesian decision theory methods seem to be widely applicable to allocating resources to evaluation studies. However, with a few exceptions these methods have not been commonly applied in evaluation nor is there much experience in using this methodology within energy efficiency evaluation. Our assessment is that these methods may have potential value for evaluation but more research and development is needed to realize this potential.³⁰⁶

Statistical power analysis is another approach to resource allocation that may be useful in some circumstances. Power analysis is a way of assessing the likelihood that a study will yield statistically significant results. The power of a statistical test is the probability of obtaining sample results that will lead to rejection of a particular null hypothesis given that the null hypothesis is false. For example, if the null hypothesis is that a program has a benefit-cost ratio of one, the statistical power is the ability of a study to demonstrate

³⁰⁵ “The Value of Improved Information: Using Decision Analysis to Quantify the Value of DSM Impact Evaluation.” (* Hildebrandt 1995). Quote on page 126.

³⁰⁶ Benefit-cost analysis is a classic case where there are several input variables each with their own sources of uncertainty. Often uncertainty in benefit-cost analysis has been handled by using scenarios analyses or sensitivity analysis. Woods and Khawaja have suggested an alternative method, multi-attribute valuation, to simplify this process and avoid assessing the propagation of uncertainty for more complex market transformation or hard to quantify programs. See “Multi-Attribute Valuation for Cost Effective Evaluation of Market Transformation and Other Hard to Quantify Programs.” (* Woods and Khawaja 2000).

that the benefit-cost ratio is significantly greater than one, given an assumed value of the true benefit-cost ratio.³⁰⁷

Propagation of Uncertainty

A closely related approach is to study how uncertainty propagates from underlying factors to the actual savings of a program or a portfolio. For example, if the total savings of a portfolio of programs is being estimated, this chapter has shown how to calculate the error bound of the total savings from the error bounds of the savings of each program. This is a simple illustration of the propagation of error, in this case from the savings of individual programs to the total savings of the portfolio of programs. In this section more complex relationships between the total savings and underlying factors are examined.

The general approach is to express the total savings as a function of two or more underlying factors. Then the amount of uncertainty in each factor is related to the uncertainty in the total savings. The results are used to assess the impact on the total savings of reducing the uncertainty in each of the factors. Finally the results are used to allocate resources to studying each factor.

This example again uses the chaining concept discussed above and examines net savings of a program, denoted NS . Net savings is calculated here as $NS = GS \times NTGR$ where GS denotes the gross savings of the program and $NTGR$ denotes the net-to-gross ratio of the program. Then, using Taylor's equation, this is approximated by:

$NS \cong E(GS) \times NTGR + GS \times E(NTGR) - E(GS) \times E(NTGR)$. With this approximation, and assuming that GS and $NTGR$ are statistically independent random variables, the standard deviation of NS is approximately equal to

$$\sqrt{E(GS)^2 \times sd(NTGR)^2 + sd(GS)^2 \times E(NTGR)^2}$$

Using the added fact that $E(NS) = E(GS) E(NTGR)$, the relative precision of NS , denoted $rp(NS)$, is approximately equal to $\sqrt{rp(GS)^2 + rp(NTGR)^2}$ as discussed in the previous section on chaining.

Table 12.9 shows how the relative precision of the gross savings and the relative precision of the net-to-gross ratio affect the relative precision of the net savings. In this case it can be seen that it is generally more effective to reduce the larger relative precision than the smaller relative precision. For example if $rp(GS) = .10$ and $rp(NTGR) = .25$, then, other considerations aside, it is better to reduce $rp(NTGR)$ to .20, giving $rp(NS) = .22$ than to reduce $rp(GS)$ to .05, giving $rp(NS) = .25$.

³⁰⁷ For an introductory discussion of statistical power analysis, see *Elements of Statistical Reasoning* (Minium and Clarke 1982). One of the classic textbooks on power analysis is *Statistical Power Analysis for the Behavioral Sciences*. (Cohen 1990).

Table 12.9: The Propagation of Error

		<i>Relative Precision of NTGR</i>				
		<i>5%</i>	<i>10%</i>	<i>15%</i>	<i>20%</i>	<i>25%</i>
<i>Relative Precision of Gross Savings</i>	<i>5%</i>	7%	11%	16%	21%	25%
	<i>10%</i>	11%	14%	18%	22%	27%
	<i>15%</i>	16%	18%	21%	25%	29%
	<i>20%</i>	21%	22%	25%	28%	32%
	<i>25%</i>	25%	27%	29%	32%	35%

An example of the use of the propagation-of-error method in energy efficiency evaluation is available in an article by Violette et al. which summarizes work performed by Xenergy for Boston Edison Company in 1992.³⁰⁸ This paper discusses the use of engineering priors within billing analysis and the variety of data sources and combination of elements involved in engineering savings estimates. The propagation-of-error method is used as a planning tool to assess where the uncertainty can be reduced most cost-effectively. In their example they compare the use of light loggers to obtain operating hours versus post-retrofit inspections to eliminate uncertainty on whether a measure is installed or not.

Another example is work performed at Wisconsin Power & Light Company by Kurt Kiefer. Partial derivatives are used to estimate the individual parameter error and then the contribution to overall uncertainty. This work was part of a evaluation planning process. The work showed the importance of good documentation of savings algorithms and data sources, providing ranges around typical point estimates for each parameter and thorough screening of existing data. This strategic work separated out the “nice to knows” from the “need to knows.”³⁰⁹

The Fixed Relative Precision Criterion

This section examines the question of how evaluation resources should be allocated to each of the programs in a portfolio of programs in order to maximize relative precision. The traditional criterion is to measure the total savings of each program with the same specified relative precision.

To analyze this approach, the following assumptions are made:

1. The objective is to estimate the savings of each program with a fixed relative precision,

³⁰⁸ “Statistically-Adjusted Engineering Estimates: What Can the Evaluation Analyst Do About the Engineering Side of the Analysis?” (Violette et al. 1993). Correcting typographical errors in Equation

$$\Delta S = \left[\sqrt{\left(\frac{\partial \Delta x}{\partial x} \right)^2 + \left(\frac{\partial \Delta y}{\partial y} \right)^2 + \left(\frac{\partial \Delta z}{\partial z} \right)^2} \right]_{x=\bar{x}, y=\bar{y}, z=\bar{z}}$$

b on page 659:

³⁰⁹ “A Framework for Strategic Evaluation Planning.” (* Kiefer 1993).

2. A efficiently stratified sample of projects will be evaluated from each program and an essentially unbiased estimate of the actual savings of each project will be determined using ratio estimation,
3. The strength of the tracking estimates available in each program is measured by a parameter called the error ratio of the program (defined in the chapter on Sampling), and
4. Each program has a relatively large number of projects.³¹⁰

Under these assumptions, the sample size that is required for each individual program can be calculated using the methods discussed in the Sampling chapter. Since the number of projects is assumed to be large, the sample size can be determined by the equation

$$n = \left(\frac{1.645 \text{ } er}{D} \right)^2$$

Here D is the desired relative precision and er denotes the assumed error ratio of each program. For example, if $D = 0.10$ and $er = 1.0$ then

$$n = \left(\frac{1.645 \times 1.0}{0.10} \right)^2 = 271$$

Table 12.10 shows an example for the four programs previously considered in Table 12.2. The third column shows the assumed value of the error ratio of each program. The column labeled Sample Size shows the required sample size calculated as above. The expected error bound for each program has been calculated as the product of the expected savings and the desired relative precision. Finally the pooling method previously discussed has been used to calculate the error bound and the expected relative precision of the total savings across all four programs.

Table 12.10: Sample Sizes for a Fixed Relative Precision of 10% at 90% Confidence

<i>Program</i>	<i>Expected Savings</i>	<i>Error Ratio</i>	<i>Desired Relative Precision</i>	<i>Sample Size</i>	<i>Expected Error Bound</i>
A	50,000	1.0	10%	271	5,000
B	50,000	0.6	10%	97	5,000
C	100,000	0.5	10%	68	10,000
D	500,000	0.9	10%	219	50,000
Total	700,000		7%	655	51,478

The equal relative precision rule may not provide a suitable allocation of evaluation resources among the programs since the sample size does not reflect the size of each program or the particular information that is needed for each individual program. For

³¹⁰ If the number of projects is small or moderate, a finite population correction can be added, as discussed in the chapter on sampling.

example, although Program A has only 7% of the total savings, it has been allocated over 40% of the total sample under this rule.

Optimal Allocation for the Overall Savings of the Portfolio

A different criterion is to allocate the sample size to each program to estimate the total savings of the portfolio with the best possible statistical precision. To explain this approach, the preceding assumptions are replaced with the following:

1. The objective is to estimate the total evaluation savings of the entire portfolio with the smallest possible error bound,
2. A sample of projects will be evaluated from each program and an unbiased estimate of the actual savings of each project will be determined,
3. The cost of evaluating each sample project is about the same across the projects of the portfolio (In practice this implies that similar methods are being used in each program of the portfolio),
4. The strength of the tracking estimates available in each program is measured by a parameter called the error ratio of the program (defined in the Sampling chapter), and
5. Each program has a relatively large number of projects.

Under these assumptions the overall sample size should be allocated to each program in proportion to the product of the expected savings of each program and the error ratio of the program. Table 12.11 provides an example. To make the new results comparable to the preceding example shown in Table 12.10, it has been assumed that a total sample of 655 projects is to be used across all four programs.

As before, the third column shows the error ratio, which measures the uncertainty in the relationship between actual savings and the tracking estimate of savings. The column labeled “Allocator” is the product of the expected savings of each program and the error ratio. In the column labeled “Optimal Sample Size,” the calculations for the sample size that should be allocated to each program are provided. In the case of program A, for

example, the sample size is calculated as $\left(\frac{50,000}{580,000} \right) 655 = 56$ projects.

Table 12.11: Optimal Allocation for the Overall Precision

<i>Program</i>	<i>Expected Savings</i>	<i>Error Ratio</i>	<i>Allocator</i>	<i>Optimal Sample Size</i>	<i>Expected Relative Precision</i>	<i>Expected Error Bound</i>
A	50,000	1.0	50,000	56	22%	10,946
B	50,000	0.6	30,000	34	17%	8,479
C	100,000	0.5	50,000	56	11%	10,946
D	500,000	0.9	450,000	508	7%	32,837
Total	700,000		580,000	655	5%	37,280

Once the sample size has been specified for each program, the expected relative precision can be calculated from the equation $rp = 1.645 \frac{er}{\sqrt{n}}$ assuming that there are a relatively large number of projects in each program, as discussed in the chapter on Sampling. Then, the expected error bound associated with each program can be calculated as the product of the expected savings and the expected relative precision, as shown in the last column of Table 12.11. The final step is to calculate the expected error bound and relative precision of the set of four programs taken together, as was discussed in the section on pooling. If it seems desirable to change the overall error bound or relative precision, the total sample size, 655, can be increased or decreased until the results seem appropriate. Finally, the funds allocated to each program can be calculated from the sample size, assuming that the unit cost per sample point is known.

Comparing Table 12.10 to Table 12.11, it can be seen that in this example without increasing the total sample of 655 projects, the statistical precision associated with the overall savings can be improved from $\pm 7\%$ to $\pm 5\%$ using the second criterion.

The allocations of the total sample are very different in Table 12.10 and Table 12.11. In Table 12.10 the fraction of the total sample that is allocated to each program is proportional to the square of the error ratio, without regard to the expected savings of the program. In Table 12.11, by contrast, the fraction of the total sample that is allocated to each program is proportional to the product of the expected savings and the error ratio of the program. Under the latter approach, the size of the program and the uncertainty of savings both have an important effect on the allocation of the evaluation resources.

The preceding methods can be extended to relax the assumption that the cost of evaluating each sample project is about the same across the projects of the portfolio. The first step is to assess the variable cost per project for each program. Then redefine the allocator to be the product of the expected savings of each program and the error ratio, divided by the square root of the unit cost of each program. The new allocator is used to divide up the total assumed sample size across all programs. The total assumed sample size can be adjusted until the total allocation costs are acceptable.

Table 12.12: Optimal Allocation with Different Unit Costs in each Program

<i>Program</i>	<i>Expected Savings</i>	<i>Error Ratio</i>	<i>Fixed Eval. Cost</i>	<i>Unit Cost / Project</i>	<i>Allocator</i>	<i>Opt. Sam. Size</i>	<i>Exp. Rel. Prec.</i>	<i>Exp. Error Bound</i>	<i>Exp. Eval. Cost</i>
A	\$50,000	1.0	\$30,000	\$100	5,000	138	14%	7,013	\$43,754
B	\$50,000	0.6	\$30,000	\$100	3,000	83	11%	5,432	\$38,252
C	\$100,000	0.5	\$100,000	\$1,000	1,581	43	12%	12,472	\$143,494
D	\$500,000	0.9	\$100,000	\$1,000	14,230	391	7%	37,415	\$491,444
Total	\$700,000				23,811	655	6%	40,424	\$716,944

Table 12.12 shows an example. Two additional columns show the fixed cost of evaluation and the unit evaluation cost per sample project for each program. Then the allocator is calculated as the product of the expected savings and the assumed error ratio divided by the square root of the unit evaluation cost per project. For example, for program A, the allocator is $\frac{50,000 \times 1.0}{\sqrt{100}} = 5,000$. This provides an optimal sample size

for program A as $\left(\frac{5,000}{23,811} \right) 655 = 138$. Then the expected relative precision for program

A is calculated as $1.645 \frac{1.0}{\sqrt{138}} = 14\%$ and the expected error bound as

$14\% \times 50,000 = 7,013$. The expected evaluation cost for program A then becomes: $\$30,000 + (138 \times \$100) = \$43,754$. The expected error bound for the total savings of all four programs was calculated as

$$\sqrt{7,013^2 + 5,432^2 + 12,472^2 + 37,415^2} = 40,424.$$

The expected relative precision for the total savings of all four programs was calculated as $\frac{40,424}{700,000} = 6\%$. In summary, the last column of shows the optimal way to allocate a

total evaluation expenditure of \$716,944 among the four programs under the assumptions given in the first four columns. The total assumed sample of 655 projects can be increased or decreased if desired, to fit the available budget and to provide the desired overall statistical precision.

The preceding method takes into account the following four factors:

- The amount of savings expected from each program,
- The uncertainty about the savings,
- The cost of evaluating each sample project in the program,

- The variability of savings in the population of projects in the program, as measured by the coefficient of variation or the error ratio.

The planner may also want to consider other factors such as:

- Whether the program is expected to grow or shrink in the future, and
- How long it has been since the last evaluation and how much the program has changed in the interim.

In considering factors such as these, the planner may want to adjust the optimal sample sizes shown in Table 12.12 for one or more of the programs. The adjusted sample sizes can be used to recalculate the expected relative precision, error bound and evaluation cost for each program and across all programs. By working interactively with a spreadsheet similar to Table 12.12, a suitable plan can be developed.³¹¹

³¹¹ David Jump, Devan Johnson and Linda Farinaccio have developed a tool for assessing several sources of uncertainty, analyzing the propagation of uncertainty, and selecting the most cost-effective M&V plan option. The authors observe that more expensive M&V plans might yield less risk. See “A Tool to Help Develop Cost-Effective M&V Plans.” (* Jump et al. 2000).

Chapter 13: Sampling

Introduction and Key Issues

All evaluation studies collect data from participants, non-participants, or the market to provide information for evaluation analysis. Unless all relevant members of a group have data collected from them (a census), some type of sampling is used. This means that sampling is used in most evaluation studies.

After the introduction, the chapter is divided into three major parts. The first part provides the basic statistical background involved in simple random sampling. This material can be found in most graduate statistics textbooks.³¹² The second part explains the basic ideas and methods of stratified ratio estimation. Most of this material is based on relatively recent literature in finite population sampling and will not be found in most statistics textbooks. However, these methods have been used extensively in impact evaluation studies and have proven to be effective.^{313, 314} The third part of the chapter is the sampling roadmap itself. The sampling roadmap describes one recommended approach to statistical sample design and analysis in a typical evaluation study. The roadmap builds on the terminology and methods of the first two sections but discusses the specific chronological steps that are recommended.

The most important section of this chapter for program implementers is the discussion of the population database. This section raises issues that are best addressed as part of program design and the design of the tracking database. A general understanding of the ratio model and the error ratio in characterizing the accuracy of the tracking estimates of savings could also significantly benefit program designers. They may also obtain some value from an overview of the sampling roadmap.

Many evaluation planners and study directors could find the background provided in the first two parts informative. The roadmap provides a straightforward guide for evaluators and policy makers.

³¹² A classic text is *Social Statistics* by Hubert M. Blalock, Jr., Stouffer Award recipient from the American Sociological Association for his outstanding contributions in social research and research methodology. (Blalock 1979)

³¹³ Much of the underlying theory was published in "Finite Population Sampling with Multivariate Auxiliary Information," (Wright 1983). These methods were used to develop the 1991 impact evaluation plan for Western Massachusetts Electric Company. Wright provided supporting testimony before the Department of Public Utilities of the Commonwealth of Massachusetts in DPU 91-44 on April 19, 1991.

³¹⁴ Some recent studies using these methods in California are *SCE Non-Residential New Construction Persistence Study*. (RLW Analytics 1998); *Impact Evaluation of PG&E & SCE's 1994 NonResidential New Construction Programs*. (RLW Analytics 1996); *Southern California Edison 1998 Non-Residential New Construction Evaluation*. (RLW Analytics et al. 1999); and *Final Report - 1999-2001 Building Efficiency Assessment (BEA) Study: An Evaluation of the Savings By Design Program*. (RLW Analytics 2003). All of these reports can be found on the CALMAC web site at <<http://www.calmac.org/>>.

Skills Required for Sampling

The population database work and simple random sampling (or a census) does not require an advanced statistics background. Stratified ratio estimation is somewhat more complex but builds on the more familiar concepts and techniques of simple random sampling. The sampling roadmap provides a step-by-step approach. Yet, it probably still requires someone to have basic training and/or experience in statistics to ensure that it is understood and applied correctly. Given this, persons conducting and reviewing this work would need to have at least basic graduate statistics or equivalent experience with a mentor in this area.

Key Issues

In a typical evaluation study, data are collected and analysis is conducted for a sample of units, usually projects or customers, selected from a given population. By following statistical sampling methods, the data collection and analysis usually can be limited to a relatively small sample. For example, in an impact evaluation study, project-specific measurement and verification (M&V) analysis might be carried out for a sample of 50 to 300 projects selected from the 1,500 projects implemented in the program in a given year.

When it is used effectively, sampling can improve the overall quality of an evaluation study. By limiting resource-intensive data collection and analysis to a relatively small fraction of all projects, more attention can be devoted to each sample project. The goal of the sampling and research design of an impact evaluation, for example, is a sound, defensible, unbiased determination of the actual gross and net savings for the overall program. Some measurement error is acceptable for each sample project – especially if the measurement error is small relative to the sampling variability. But measurement bias should be minimized since it will propagate through the analysis. It is equally important to minimize other sources of bias such as non-response, self-selection, and deliberate substitution.³¹⁵

If the sample projects are selected following an efficient sample design, and if the data collection and site-specific analysis is free of substantial bias, then the statistical analysis can provide an essentially unbiased estimate of each population characteristic of interest and a good measure of the achieved statistical precision. The statistical analysis can also provide measures of population variability to guide future evaluation studies.

If a suitable sample design is followed in selecting the sample projects, the estimates of population characteristics developed from the sample data will generally be close to the true values that would have resulted if the same data collection and project-specific M&V analysis had been carried out for all projects in the population. Moreover, an error bound can be calculated to assess the statistical precision of the results.

³¹⁵ See Chapter 12 on Uncertainty for a further discussion of potential bias issues that need to be addressed within the evaluation planning process and in analysis and reporting.

However, the estimates calculated from the sample may be seriously biased and the statistical error bound may be misleading if (a) the final sample is substantially different than the primary sample due to non-response, refusals, or substitutions or (b) the research design, data collection, or analysis is seriously biased. Therefore it is sound practice to reinvest some of the resources saved by sampling into efforts to (a) promote high response rates and minimize substitutions, and (b) minimize bias in the data collection, research design, and analysis.

General Steps in a Study

This chapter summarizes the sampling methodology that is generally followed in an evaluation study. As shown in Figure 13.1, a study usually moves through the following seven sequential stages:

1. **Objectives** – identify the objectives of the study,
2. **Population Database** – create the list of all units in the target population,
3. **Data Collection Approach** – plan the basic methodology that will be used to develop the desired information for the sample units,
4. **Sample Design** – develop the approach to selecting the sample units,
5. **Sample Selection** – select the sample units following the sample design,
6. **Data Collection and M&V** – develop the desired information for each sample unit,
7. **Statistical Analysis** – extrapolate the sample data up to the population, calculate error bounds, and develop the information needed to guide future sample designs, and
8. **Final Report** – summarize the methodology, findings and recommendations of the study.

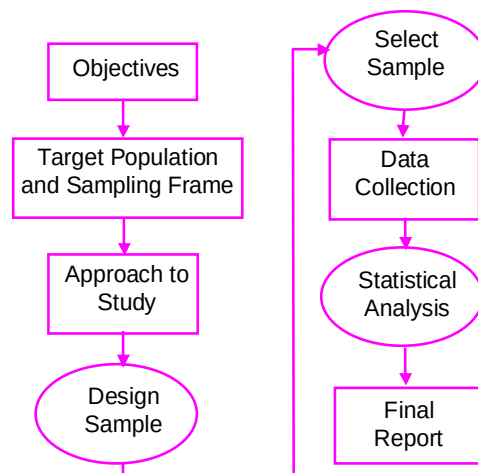


Figure 13.1: Stages in a Study

However, the conceptual foundations cannot be discussed effectively in this linear fashion. For example, the sample design and statistical analysis methodologies have to be considered together as a coherent strategy for developing information. Moreover, it is easier to develop a thorough conceptual understanding of the relatively simple approach

of simple random sampling before embarking on the more complex methodology of stratified ratio estimation. Therefore the conceptual discussion given in the first two parts does not strictly follow this sequence shown in Figure 13.1. This chapter will, however, return to this chronological sequence in the sampling roadmap given in the final section of this chapter.

Population Database

The starting point of any discussion of finite population sampling is the target population itself. For example, in an impact evaluation study, the population may be the set of projects implemented in the program in a given year. In a baseline study the population may be the set of all non-residential customers served by a given utility or the set of all vendors of a given appliance in a specified region.

The population database is a list of each unit in the population, with relevant information for each unit. If the target population is a set of projects implemented in a particular program in a given year, then the population database is a list of each of these projects, together with supporting information for each project, developed from the program tracking system.

In the context of simple random sampling, the minimum requirement is the information needed to identify each project so that the desired data can be collected for the projects that fall into the sample. The population database generally should also have a suitable measure of the size of each unit in the population. This usually is either the tracking estimate of the savings of each project in the program, or the annual energy usage of each unit in the population. For more information about constructing the population database, please see the sampling roadmap given at the end of this chapter.

Simple Random Sampling

In simple random sampling, a sample of a given size, denoted n , is selected from the projects in the population following any randomized procedure in which all possible subsets of n projects are equally likely to be selected. For example, assume there is a complete computerized list of all of the projects in the population.³¹⁶ This is called a sampling frame. A simple random sample is selected by assigning a random number to each project, sorting the sampling frame according to the random number, and designating the first n projects in the sorted list to be the sample.

In evaluation, simple random sampling can be very effective if the projects in the population do not vary too much in size. This may be the case for some residential programs and certain non-residential programs if larger projects are suitably divided into

³¹⁶ The population must be unambiguously defined. Although most programs are ongoing, it is not appropriate to allow the population to be constantly changing. Therefore it is common to define the population for an evaluation study to be the set of projects paid in a given program in a given year.

subprojects that can be independently evaluated. The statistical theory of simple random sampling is also of interest because it is the foundation of all other methods of sampling.

Basic Definitions

Some technical terminology and notation is needed to discuss the methodology of simple random sampling. This notation will also be useful in discussing the more complex sampling methods to follow. Let N denote the number of projects in the population and assume that the projects are labeled $i = 1, \dots, N$. Let y denote any measurable variable of interest, such as gross or net savings. Then let y_i denote the value of y for project i .

Y will denote the true total of y for all N projects in the population, i.e., $Y = \sum_{i=1}^N y_i$. The population mean of y is denoted as μ :

$$\mu = \frac{Y}{N} = \frac{1}{N} \sum_{i=1}^N y_i.$$

The population variance of y is denoted by σ^2 :

$$\sigma^2 = \frac{1}{(N-1)} \sum_{i=1}^N (y_i - \mu)^2.$$

The population standard deviation of y is $\sigma = \sqrt{\sigma^2}$ and the population coefficient of variation of y is $cv = \sigma/\mu$. These are all parameters of the population; in practice they are generally unknown.

Now assume that a random sample of n projects has been selected and that the sample projects are labeled as $i = 1, \dots, n$. Assume that the value of y_i has been observed^{317, 318}

³¹⁷ In discussing sampling, it is common to neglect measurement error. In the context of impact evaluation, for example, the population mean μ is regarded to be the mean of the values of y_i that would be hypothetically observed if all projects in the population were subjected to the same data collection and analysis techniques planned for the sample projects. Only minor changes in our discussion would be needed to accommodate random measurement error. Most of our results would be unchanged with the exception of the finite population correction factor.

³¹⁸ Measurement bias is much more serious than random measurement error. If the planned data collection and analysis techniques give a biased estimate of the true savings of each project that is analyzed, then μ itself will be biased in the sense that it will not be equal to the population mean of the true savings of all projects. In this case, although the procedures discussed in this chapter will yield an unbiased estimate of μ with measurable statistical precision, it is generally very difficult to estimate the difference between μ and the population mean of the true savings of all projects. See the chapter on Uncertainty for a more complete discussion of bias.

for each sample project i .³¹⁹ Corresponding to the population parameters that have been defined above, the sample statistics are defined as:

- The sample mean of y is $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$,
- The sample variance of y is $s^2 = \frac{1}{(n-1)} \sum_{i=1}^n (y_i - \bar{y})^2$,
- The sample standard deviation of y is $s = \sqrt{s^2}$, and
- The sample coefficient of variation of y is $\hat{c}v = s/\bar{y}$.

In the context of simple random sampling, each of these sample statistics can be regarded as an estimator of the corresponding population parameter. The key issue is the performance of statistics as estimators. This issue bears on the choice of sample size and the statistical precision to be expected of an estimator. There is generally particular interest in the statistical precision of the sample mean $\bar{y}$ as an estimator of the population mean μ .³²⁰ The central concept needed to address these issues is the sampling distribution of $\bar{y}$.

Sampling Distribution of the Sample Mean

The preceding section discussed the procedure of selecting a random sample of a fixed size from a fixed population, observing the value of a particular variable y for each sample project, and calculating the sample mean, $\bar{y}$. Now consider the following conceptual experiment. Suppose the procedure just described is repeated a large number of times. The distribution of values of $\bar{y}$ that would be observed in repeated sampling is called the sampling distribution of the sample mean.³²¹

The sampling distribution can be used to define the expected value, variance, and standard deviation of $\bar{y}$. Under simple random sampling, the following results are known:

1. The expected value of $\bar{y}$ in repeated sampling is equal to μ , the population mean of y . In other words, $\bar{y}$ is an unbiased estimator of μ .

³¹⁹ One form of measurement error can arise when some of the sample projects are so large that only part of the total space or measures can be explicitly evaluated. In this situation it is important to avoid measurement bias arising from the selection of the spaces or measures to be studied.

³²⁰ Analogous methods are also readily available for a sample proportion as an estimator of a population proportion. These methods are commonly used to help plan market research and other qualitative studies, especially if the units in the population are fairly homogeneous in size.

³²¹ This idea of repeated sampling and its use in statistical inference comes from the frequency interpretation of probability that is the basis for the classical theory of statistics. The alternative Bayesian approach relies upon the subjective interpretation of probability. See *Bayesian Statistics for Evaluation Research: An Introduction* (Pollard 1986). For a brief discussion of the Bayesian approach and added references please see the chapter on Uncertainty.

2. The standard deviation of $\bar{y}$ in repeated sampling, sometimes denoted as $sd(\bar{y})$, is equal to $\frac{\sigma}{\sqrt{n}}$, provided that the population size N is large relative to the sample size n . Here σ is the population standard deviation of y . In general, $sd(\bar{y}) = \sqrt{1 - \frac{n}{N}} \frac{\sigma}{\sqrt{n}}$ where $\sqrt{1 - \frac{n}{N}}$ is called the finite population correction factor.
3. If the sample size n is large but substantially smaller than the population size N , then the sampling distribution of $\bar{y}$ can be approximated by the normal probability distribution.

All three of these results can be combined into the following statement: In repeated sampling, the probability is about 90% that the sample mean $\bar{y}$ will fall within $\pm 1.645 sd(\bar{y})$ of the population mean, μ . This key result is the basis for calculating the expected statistical precision for a specified sample size, for choosing the sample size to provide a desired statistical precision, for estimating the statistical precision from the sample data that have been collected, and for estimating the population parameters needed to plan future studies.

Expected Statistical Precision

Refer again to the key statement: In repeated sampling, the probability is about 90% that the sample mean $\bar{y}$ will fall within $\pm 1.645 sd(\bar{y})$ of the population mean μ . This statement is used to define the two most important measures of statistical precision: the error bound and the relative precision. The quantity $1.645 sd(\bar{y})$ is called the expected error bound of $\bar{y}$ as an estimator of μ at the 90% level of confidence.³²² This is denoted as $eb(\bar{y})$. Assuming that μ is greater than zero, the quantity $eb(\bar{y})/\mu$ is called the expected relative precision at the 90% level of confidence and is denoted as rp .

Using result 2 from the preceding section, the expected error bound of $\bar{y}$ is given by the equation $eb(\bar{y}) = 1.645 \frac{\sigma}{\sqrt{n}}$ provided that the population size N is large relative to the sample size n . In general, $eb(\bar{y}) = 1.645 \sqrt{1 - \frac{n}{N}} \frac{\sigma}{\sqrt{n}}$. Similarly, the relative precision is

³²² The 90% level of confidence has become conventional in evaluation work. The authors recommend that the 90% level of confidence be used consistently to avoid needless confusion.

given by the equation $rp = 1.645 \frac{cv}{\sqrt{n}}$ if N is large, and $rp = 1.645 \sqrt{1 - \frac{n}{N}} \frac{cv}{\sqrt{n}}$ in general.³²³

These results show that the expected statistical precision depends on the following three factors:

1. The sample size n ,
2. The population size N , especially if the latter is relatively small, and
3. The variability of y in the population, measured either by σ or by cv .

Choice of Sample Size

The preceding results can be used to guide the choice of the sample size. The analysis requires an assumption concerning the variability of y in the population, measured by cv . In many applications, the cv of y can be estimated adequately by calculating the coefficient of variation of the measure of size that is available in the population database. Of course the number of units in the population N is also known from the population database.

Given N and the estimated value of cv , several techniques provide guidance for the choice of sample size. One approach is to create a table showing the expected relative precision for different choices of sample size n , using the equation

$$rp = 1.645 \sqrt{1 - \frac{n}{N}} \frac{cv}{\sqrt{n}}.$$

Another approach is to set the desired relative precision and calculate the required sample size. The desired relative precision is usually denoted D . Assume initially that the population size N is large relative to the sample size n . Using the equation

$$rp = 1.645 \frac{cv}{\sqrt{n}},$$

the required sample size can be estimated as a function of the population cv :

$$n = \left(\frac{1.645 cv}{D} \right)^2.$$

³²³ In evaluation, the focus is sometimes on the population total $Y = \sum_{i=1}^N y_i$. When simple random sampling is used, the population total is estimated using the statistic $\hat{Y} = N \bar{y}$. Since N is not subject to sampling variability, the standard deviation of $\hat{Y}$, denoted $sd(\hat{Y})$, is equal to $N sd(\bar{y})$. As usual, the error bound of $\hat{Y}$ is $eb(\hat{Y}) = 1.645 sd(\hat{Y}) = N eb(\bar{y})$. The expected relative precision of $\hat{Y}$ is identical to the expected relative precision of $\bar{y}$.

For example, if $D = 0.10$ and assuming $cv = 0.75$, then the preceding equation gives

$$n = \left(\frac{1.645 \times 0.75}{0.10} \right)^2 = 152 .$$

Table 13.1 shows the results of this type of calculation for various values of cv and D . The table shows clearly that both of these parameters have a dramatic impact on the required sample size. It will be shown later in this chapter that the relevant population variability can often be reduced very substantially by using stratified ratio estimation instead of simple random sampling.

Table 13.1: Required Sample Size Assuming a Large Population

		<i>Coefficient of Variation, cv</i>						
		<i>0.5</i>	<i>0.75</i>	<i>1.0</i>	<i>1.5</i>	<i>2.0</i>	<i>2.5</i>	<i>3.0</i>
<i>Desired Relative Precision D</i>	<i>0.25</i>	11	24	43	97	173	271	390
	<i>0.20</i>	17	38	68	152	271	423	609
	<i>0.15</i>	30	68	120	271	481	752	1,082
	<i>0.10</i>	68	152	271	609	1,082	1,691	2,435

If there are a small or moderate number of projects in the population, then a finite population correction can be applied to the preceding method. The first step is to use the preceding equation to calculate an initial, unadjusted estimate of the required sample size, denoted n_0 :

$$n_0 = \left(\frac{1.645 \, cv}{D} \right)^2 .$$

Then the following finite population correction is applied:

$$n = \frac{n_0}{1 + n_0/N} .$$

For example, if the population has 500 projects so that $N = 500$, and as before $D = 0.10$ and $cv = 0.75$, then the required sample size can be calculated in two steps as follows:

$$n_0 = \left(\frac{1.645 \times 0.75}{0.10} \right)^2 = 152 , \text{ and}$$

$$n = \frac{152}{1 + 152/500} = 117 .$$

After the finite population adjustment, the sampling fraction is $\frac{n}{N} = \frac{117}{500} = 0.23$.

Table 13.2 illustrates the second step of this calculation for different population sizes, assuming $n_0 = 152$. The table illustrates an important fact: if the population is large, then the finite population correction has little effect on the sample size n , but if the population is small, then the finite population correction is important..

Table 13.2: Finite Population Adjustment for Different N

N	100	500	1,000	5,000	10,000
n	60	117	132	148	150
n/N	0.60	0.23	0.13	0.03	0.01

The table illustrates a second important fact that has important policy implications. If the population is large, then the sample size usually is a small fraction of the population but if the population is small, the sample can be a substantial portion of the population. This can have an important impact on the cost of evaluating small programs. In other words, the traditional criterion of obtaining $\pm 10\%$ relative precision at the 90% level of confidence for each program would impose a disproportionately larger burden on small programs than large programs.³²⁴

In place of this traditional criterion, the desired relative precision can be relaxed for smaller programs. Table 13.1 shows that this can reduce the required sample size substantially. Alternatively, the authors recommend that the evaluation budget allocated to each program be guided by the resource allocation principles discussed in Chapter 12 on uncertainty.³²⁵

Statistical Analysis

It is useful to summarize some of the results discussed so far. It has been shown that in repeated sampling, the probability is about 90% that the sample mean $\bar{y}$ will fall within $\pm 1.645 \text{ } sd(\bar{y})$ of the population mean μ . The quantity $1.645 \text{ } sd(\bar{y})$ has been called the expected error bound at the 90% level of confidence and is denoted $eb(\bar{y})$. The quantity

³²⁴ The original CADMAC Protocols specified that if the program population was small, a census should be attempted. For example, if a program had 100 projects, an attempt was made to evaluate each one of the 100 projects. Suppose that a number of projects were dropped from the evaluation, leaving a convenience sample or self-selected sample of perhaps 60 projects. A casual reading of Table 13.2 might suggest that this would be adequate in this example since 60 is the required sample size under the $\pm 10\%$ relative precision criterion. However, while a random sample of 60 projects will provide an unbiased estimate of the total savings of all 100 projects with measurable precision, a convenience or self-selected sample is vulnerable to bias that is difficult to quantify.

³²⁵ The resource allocation approach tends to focus on the expected error bound of the estimator of the population total, $\hat{Y}$, denoted $eb(\hat{Y})$. Since $eb(\hat{Y}) = rp \hat{Y}$, the expected error bound of the total can always be calculated by multiplying the expected relative precision by the expected value of $\hat{Y}$.

$rp = eb(\bar{y})/\mu$ has been called the expected relative precision at the 90% level of confidence. Under simple random sampling, recall that $sd(\bar{y}) = \frac{\sigma}{\sqrt{n}}$ if the population size N is large relative to the sample size n , and in general $sd(\bar{y}) = \sqrt{1 - \frac{n}{N}} \frac{\sigma}{\sqrt{n}}$.

Now assume that a specific random sample has been selected of n projects labeled as $i = 1, \dots, n$, and that the value of y_i has been observed for each sample project i . The sample data is used to calculate the following sample statistics: the sample mean of y , $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$, the sample variance of y , $s^2 = \frac{1}{(n-1)} \sum_{i=1}^n (y_i - \bar{y})^2$, the sample standard deviation of y , $s = \sqrt{s^2}$, and the sample coefficient of variation of y , $\hat{cv} = s/\bar{y}$. Here $\bar{y}$ is an estimator of the population mean μ , s is an estimator of the population standard deviation σ , and $\hat{cv}$ is an estimator of the population coefficient of variation cv .

These statistics, in turn, can be used with the equations just given to estimate the standard deviation of $\bar{y}$, as well as the corresponding error bound and relative precision. When the standard deviation of a statistic has been estimated, the result is called a standard

error. The standard error of $\bar{y}$ is $se(\bar{y}) = \frac{s}{\sqrt{n}}$ if the population size N is large relative to

the sample size n , and $se(\bar{y}) = \sqrt{1 - \frac{n}{N}} \frac{s}{\sqrt{n}}$ in general.³²⁶ Then the error bound can be calculated as $eb(\bar{y}) = 1.645 se(\bar{y})$ and the relative precision as $rp = eb(\bar{y})/\bar{y}$.

The 90% confidence interval for μ is defined to be $\bar{y} \pm eb(\bar{y})$. In the hypothetical context of repeated sampling, the probability is about 90% that the confidence interval (which will vary from sample to sample) will contain the true value of μ .³²⁷

Applicability to Impact Evaluation

In impact evaluation, an important objective usually is to assess the true annual energy savings of all projects undertaken in a program. One way to do this is to select a random sample of projects from the population of all projects and then assess the annual energy savings of each sample project using suitable data collection and analysis techniques. Under the assumptions of simple random sampling, the average of the annual energy savings of the sample projects is an unbiased estimator of the average annual energy

³²⁶ If the population total is of interest, the standard error of $\hat{Y}$ is equal to N times the standard error of $\bar{y}$, i.e., $se(Y) = N se(\bar{y})$.

³²⁷ Sometimes it is useful to calculate a one-sided confidence interval. For example, a one-sided confidence interval could be used to state that the savings are 'no smaller than ...'.

savings of all projects in the population from which the sample has been drawn. Thus far this chapter has described procedures for assessing the statistical precision of this estimator and results that can guide the choice of the sample size.

The assumptions of simple random sampling are the key to these results. It is assumed that the sample is randomly selected from the entire target population and that the actual annual savings are observed for all projects in the sample. So the possibility of participant refusal or other difficulties has been excluded that would lead to the replacement of a sample project by an alternate. The methods here also assume that the actual annual savings is observed for each sample project with little or no measurement error or bias compared to the expected relative precision of the results.

If, contrary to these assumptions, there is substantial bias, perhaps due to self-selection, non-response, deliberate substitution of sample projects, or measurement bias, then the methods presented here can be seriously misleading. For example it is misleading and counterproductive to report that the average savings has been estimated with a relative precision of $\pm 10\%$ at the 90% level of confidence if there is a serious risk that the results might be in error by 25% due to bias. Bias is particularly insidious because it usually is extremely difficult to assess its magnitude and often even its direction.

Consequently, in a sound evaluation study, every practical effort should be made to minimize all of the likely sources of bias such as non-response, self-selection, deliberate substitution, and measurement bias.³²⁸ One of the important goals of sampling is to limit the sample size so that there will be adequate resources for sound data collection and analysis.

The statistical precision provided by a simple random sample is strongly influenced by the variability of annual savings from project to project in the target population. For example, Table 13.1 shows that if the coefficient of variation is three, if the desired relative precision is $\pm 15\%$, and if the population is large, then a sample size of 1,082 is needed with simple random sampling. A sample this large would usually be infeasible.³²⁹ However, this does *not* imply that statistical sampling is impractical whenever there is large variability of savings in the target population – only that simple random sampling is impractical. The stratified random sampling techniques discussed in the next section generally address this situation very effectively, especially if the program provides good tracking information for each project.

The sample size is also affected by the size of the target population. Assuming that the sample size is chosen to provide a fixed relative precision, the sample will generally be a small fraction of the population if the population is large, but it will be a larger fraction of the population if the population is small. Thus sampling will be inherently more cost-effective for large programs than for small programs.

³²⁸ See the chapter on Uncertainty for a more detailed discussion of potential bias, and the need to assess and report potential bias and threats to the evaluation's validity.

³²⁹ Moreover a very large sample generally precludes the steps needed to mitigate bias.

For this reason, if there are a number of small programs, it may be desirable to consolidate the programs for the purpose of evaluation. In this context, the population would be taken to be the entire set of projects undertaken in a portfolio of programs. Using the methods of this section, a simple random sample would be selected from the population of all projects in the portfolio of programs, the annual savings of each sample project would be assessed using appropriate data collection and analysis techniques, and the sample results would be used to estimate the average savings per project for all projects in the portfolio. If there are a sufficiently large number of projects within the portfolio, sampling will generally be cost-effective and there should be adequate resources for sound data collection and analysis.³³⁰

Under this approach, the consolidated sample data can be used to provide an unbiased estimate of the average savings per project for each individual program. However, the average savings for the individual programs would generally have poorer statistical precision than the average savings for the portfolio since the individual program results would be based on the sample projects that were drawn from that program. Moreover, using the simple random sampling approach, the number of sample projects from each program would not be controlled but would vary randomly from one possible sample to another. Nevertheless, the information from the consolidated impact evaluation, in conjunction with other types of information, may still provide adequate information about each program.

Stratified Ratio Estimation

Stratified ratio estimation combines a stratified sample design with a ratio estimator.³³¹ Both stratification and ratio estimation take advantage of supporting information available for each project in the population. As an example, suppose that an impact evaluation study is being undertaken to assess the annual energy savings of the projects undertaken in a given program. Suppose that the program tracking system provides an estimate of the annual energy savings of each project in the population. Suppose, furthermore, that a substantial fraction of the projects have comparatively small tracking savings but a relatively small number of projects have very large tracking savings. In this case, the coefficient of variation of the tracking savings will often be quite large, e.g., three or larger, and it can be expected that the population coefficient of variation of the

³³⁰ Consolidation will usually increase the coefficient of variation of annual savings if the average savings per project is different across the individual programs. Under simple random sampling, the larger coefficient of variation could compromise the statistical precision. These problems can usually be addressed through appropriate use of stratified ratio estimation, to be discussed in the next section.

³³¹ Statisticians have developed many other approaches to sample design and estimation, including sequential sampling, cluster sampling, multi-stage sampling, stratified sampling with mean per unit estimation, stratified sampling with regression estimation, etc. See, for example, *Sampling Techniques* (* Cochran 1977). Any of these methods may be appropriate in a particular application. The authors have found that stratified ratio estimation is generally effective in both impact and process evaluation studies, especially when (a) there is substantial variation in the size of projects in the program, and (b) the tracking system provides fairly accurate estimates of the savings of each project. These conditions are frequently true for energy conservation programs.

actual savings is also large. In this case, the simple random sampling methods described in the preceding section would not be practical.

This problem can be partly mitigated by using the tracking estimate of savings as a stratification variable. Stratifying by the tracking savings generally reduces the coefficient of variation of actual savings in each stratum thereby improving the statistical precision.³³² Moreover, the sampling fraction can be varied from stratum to stratum to further improve the statistical precision. In particular, a relatively small sample can be selected from the projects with small tracking savings, but the sample can be forced to include a higher proportion of the projects with larger tracking savings. In particular, the largest projects can, if desirable, be included in the sample with certainty.

The tracking estimates of savings can also be used in ratio estimation. In impact evaluation, one ratio of interest is the realization rate, i.e., the ratio between the total gross annual savings of all projects in the population and the total tracking savings.³³³ To understand the potential advantage of ratio estimation, suppose hypothetically that the actual savings of each project in the population is directly proportional to the savings recorded in the tracking system as illustrated in Figure 13.2.

In the extreme example illustrated in Figure 13.2, the actual savings of each project is 0.8 times the tracking estimate of savings. In other words, the tracking system systematically overstates the saving of each project by 20%. The realization rate, 0.8, is the slope of the line relating the actual savings to the tracking for every project. If the realization rate is known, then the true savings of all projects can be accurately estimated by multiplying the total tracking savings by the realization rate. Moreover, in this extreme case, the realization rate can be assessed perfectly by measuring the actual savings of any one project in the population.

³³² In this case, however, the coefficient of variation of tracking savings within each stratum usually does not provide a meaningful estimate of the coefficient of variation of actual savings within each stratum. Therefore added information is needed to estimate the expected statistical precision and to choose the sample size, e.g., from a prior sample or from a model characterizing the relationship between tracking and actual savings.

³³³ The net-to-gross ratio is another ratio of interest. Our experience has been that ratio estimation can be used to estimate essentially all parameters of interest in evaluation.

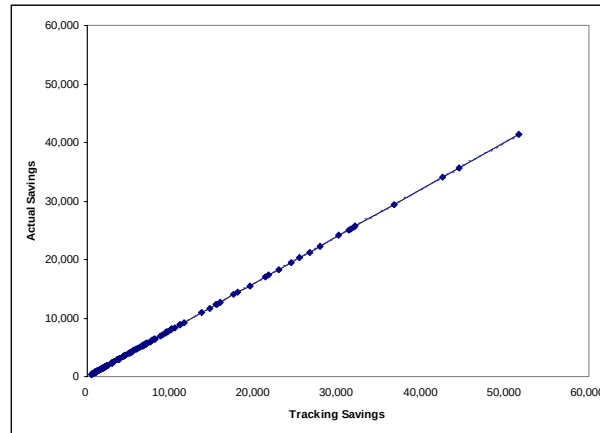


Figure 13.2: Ideal Case for Ratio Estimation

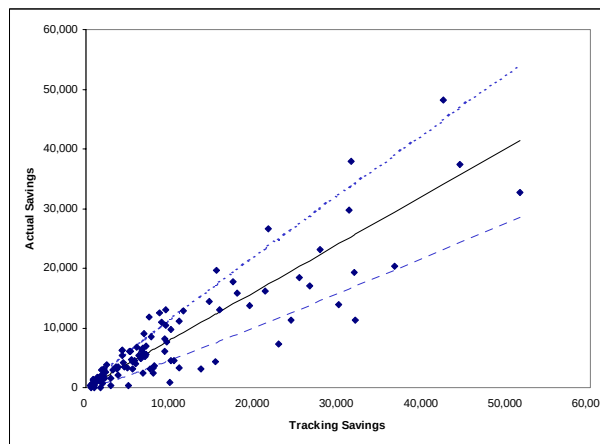


Figure 13.3: More Typical Relationship between the Actual and Tracking Savings

In practice, of course, there is always some random error in the association between the actual and tracking savings. Figure 13.3 illustrates a more typical situation. In this case the tracking estimate of savings is a good but not perfect predictor of the actual savings of each project. Nevertheless, the statistical precision can be greatly improved by using stratified ratio estimation to estimate the realization rate rather than by using simple random sampling to assess the average actual savings as discussed in the prior section.

Goals of the Section and Basic Definitions

This section will provide the tools needed to use stratified ratio estimation effectively in evaluation. The goal is to explain the underlying concepts in enough detail for users to be comfortable with the methodology. Specifically, this section will explain:

- How to estimate the population parameters of interest and to calculate the associated confidence intervals,
- How to characterize the population variation when efficiently stratified ratio estimation is to be used,

- How the expected statistical precision is related to the population variation and to the planned sample size assuming that efficient stratification is used,
- How to estimate the required sample size to achieve a desired relative precision,
- How to construct an efficiently stratified sample design, and
- How to estimate the relevant population variation from the sample for use in planning future studies.

Much of the notation needed to discuss the methodology of stratified ratio estimation is retained from the earlier discussion of simple random sampling. Let N denote the number of projects in the population and assume that the projects are labeled $i = 1, \dots, N$. Let y denote any measurable variable of interest, such as gross or net savings and let y_i denote the value of y for project i . Y denotes the true total of y for all N projects in the population, i.e., $Y = \sum_{i=1}^N y_i$, and μ_y denotes the population mean of y ,

$$\mu_y = \frac{Y}{N} = \frac{1}{N} \sum_{i=1}^N y_i.$$

Stratified ratio estimation focuses on the relationship between y and a second variable, denoted x . The value of x is assumed to be known for each project in the population,³³⁴ and to avoid minor notational inconveniences, x is assumed to be greater than zero for each project in the population. In the impact evaluation context, x is usually the tracking estimate of the savings of each project. X denotes the total of x for all N projects in the population, i.e., $X = \sum_{i=1}^N x_i$ and μ_x denotes the population mean of x ,

$$\mu_x = \frac{X}{N} = \frac{1}{N} \sum_{i=1}^N x_i.$$

The key population parameter of interest is the ratio between the population total of y and the population total of x , which is denoted B and defined by the following equation:

$$B = \frac{Y}{X} = \frac{\sum_{i=1}^N y_i}{\sum_{i=1}^N x_i}.$$

Of course, B is also equal to the ratio between μ_y and μ_x , i.e., $B = \frac{\mu_y}{\mu_x}$.

³³⁴ Stratified ratio estimation can also be used when the denominator of the ratio is unknown. For example this methodology can be used to estimate the net-to-gross ratio. In this case, a different variable, usually the measure of size in the tracking system, is used for stratification.

Stratified sample design uses knowledge about the population to add efficiency to the sample design. A stratum is any subset of the projects in the population that is based on known information. A stratification of the population is a classification of all units in the population into mutually exclusive strata that span the population. Under a stratified sample design, simple random sampling is used to select a chosen number of projects from each of the pre-established strata.

Added notation is needed to discuss stratified sampling. Let L denote the number of strata and assume that the strata are labeled $h = 1, \dots, L$. Let N_h be the total number of population projects in stratum h . Let n_h be the number of projects to be randomly selected from stratum h . Assume that n_h is greater than zero for each stratum h . Then $\sum_{h=1}^L N_h = N$, the total population size, and $\sum_{h=1}^L n_h = n$, the total sample size.

Using this notation, the stratified ratio estimator can be defined. For each project i in the sample, the case weight is defined according to the equation $w_i = N_h/n_h$ where h denotes the particular stratum that contains project i . Using the case weights, define the stratified ratio estimator of B , denoted b , as follows:³³⁵

$$b = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i x_i}.$$

The statistical precision of b can be assessed by calculating the standard error using the following equation:

$$se(b) = \frac{1}{\hat{X}} \sqrt{\sum_{i=1}^n w_i (w_i - 1) e_i^2}.$$

Here $\hat{X} = \sum_{i=1}^n w_i x_i$ and $e_i = y_i - b x_i$. Then, as usual, the error bound can be calculated as $eb(b) = 1.645 se(b)$ and the relative precision can be calculated as $rp = eb(b)/b$.

³³⁵ An equivalent equation is $b = \frac{\hat{Y}}{\hat{X}} = \frac{\sum_{h=1}^L N_h \bar{y}_h}{\sum_{h=1}^L N_h \bar{x}_h}$. Technically, the stratified ratio estimator is a biased estimator of the true population ratio. However, Cochran shows that the bias is small if the relative precision of $\sum_{h=1}^L N_h \bar{x}_h$ is small, pp. 160-167 (* Cochran 1977). In impact evaluation, the bias should be negligible if the population has been appropriately stratified by size as discussed later in this chapter.

Stratified ratio estimation can also be used to estimate the population mean or population total of y from the known population mean or population total of x . The estimator of the mean is $\hat{\mu}_y = b \mu_x$ and the corresponding standard error is $se(\hat{\mu}) = \mu_x se(b)$. The estimator of the total is $\hat{Y} = b X$ and the corresponding standard error is $se(\hat{Y}) = X se(b)$.

The Ratio Model

To develop a suitable sample design, it is necessary to characterize the relation between x and y in the population. This is done by assuming a statistical model called the ratio model. The primary equation of the ratio model is $y_i = \beta x_i + \varepsilon_i$. Here x_i and y_i denote the value of x and y for each project i in the population, β is an unknown but fixed parameter of the model that is similar to a regression coefficient, and ε_i is similar to the random error in a regression model. As in a regression model, the expected value of ε_i is assumed to be zero for each project i in the population. It is also assumed that $\varepsilon_1, \dots, \varepsilon_N$ are mutually independent. Then μ_i is defined to be the expected value of y_i given x_i . Under the ratio model $\mu_i = \beta x_i$.

Instead of assuming that the standard deviation of ε_i is constant, the standard deviation of ε_i is allowed to vary from project to project. For any project i in the population, the standard deviation of ε_i is denoted as σ_i . This is called the residual standard deviation of project i . The population error ratio of x and y , denoted er , is defined to be

$$er = \frac{\sum_{i=1}^N \sigma_i}{\sum_{i=1}^N \mu_i}.$$

The error ratio is the key measure of the population variability in the relationship between x and y for stratified ratio estimation. The role of the error ratio in stratified ratio estimation is virtually the same as the role of the coefficient of variation in simple random sampling. Figure 13.4 shows several examples of error ratios ranging from 0.4 (a relatively strong relationship) to 1.0 (a weak relationship).

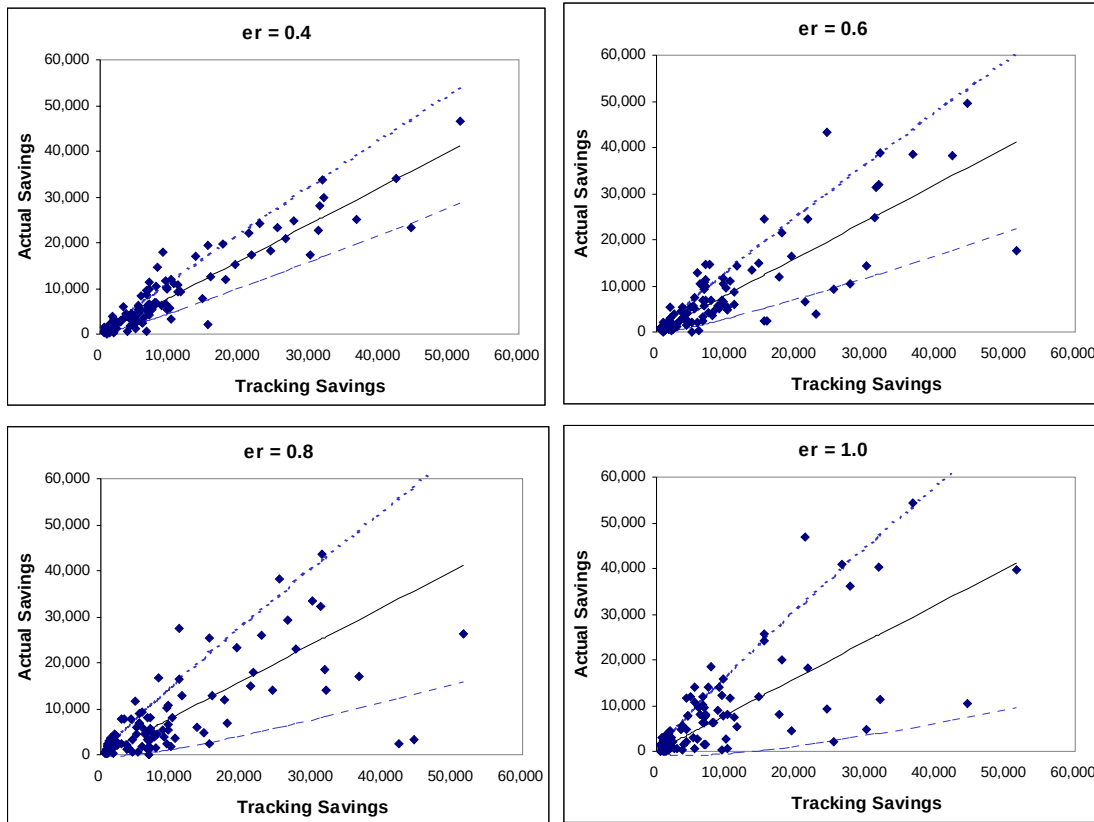


Figure 13.4: Examples of Different Error Ratios

The following specific functional form for σ_i is often assumed: $\sigma_i = \sigma_0 x_i^\gamma$. This is called the secondary equation of the model.³³⁶ The secondary equation specifies that the residual standard deviation of each project i in the population is proportional to the value of x_i raised to the power γ , pronounced gamma. A common assumption is that $\gamma = 0.8$. This specification is used in constructing efficiently stratified sample designs and to assist in the estimation of the error ratio from a prior sample.

The secondary equation includes a parameter denoted σ_0 . This parameter is determined by the error ratio as follows:

$$\sigma_0 = er \frac{\sum_{i=1}^N \mu_i}{\sum_{i=1}^N x_i^\gamma}.$$

³³⁶ Sarndal writes the secondary equation as $\sigma_i^2 = c x_i^\gamma$ (Sarndal et al. 1992), pp. 449.

Sampling Distributions

The simple random sampling section discussed the concept of repeatedly selecting a random sample of a fixed size from a fixed population, observing the value of a particular variable y for each sample project, and calculating appropriate statistics. This concept was used to define the sampling distribution of a statistic such as the sample mean. This same concept of repeated sampling is used in the present discussion with one extension. Instead of regarding y_i as fixed for each project i , y_i is assumed to vary randomly from sample to sample, generated by independent realizations of the ratio model. In other words, the sample is regarded to be randomly determined following the prescribed sample design, and the true values of y_i are considered to be randomly determined for all N units in the population following the ratio model. A more in-depth discussion of this concept can be found in Sarndal.³³⁷

Expected Statistical Precision and Choice of Sample Size

A key result for stratified ratio estimation is the following: Assuming that the ratio model is accurate, that the sample design is efficiently stratified for the model as described later in this section, that the population size N is large and that the 90% level of confidence is used, then the expected relative precision of the stratified ratio estimator is approximately equal to

$$rp = 1.645 \frac{er}{\sqrt{n}}.$$

This result can be used to guide the choice of the sample size. Suppose that the desired relative precision is denoted D . Under the preceding assumptions, the sample size needed to provide an expected relative precision of D at the 90% level of confidence is approximately

$$n = \left(\frac{1.645 er}{D} \right)^2.$$

These are the same equations given in the discussion of simple random sampling, but with the coefficient of variation replaced by the error ratio. If N is moderate or small, the finite population correction factor can be used as a first approximation as in simple random sampling. A somewhat more complex but more accurate way of adjusting the large population results for the size of the population will be presented later in this chapter.

For example, if $D = 0.10$ and $er = 0.5$, then the preceding equation gives

³³⁷ (Sarndal et al. 1992), pages 448-471.

$$n = \left(\frac{1.645 \times 0.5}{0.10} \right)^2 = 68 .$$

Table 11.3 shows the results of this type of calculation for various values of er and D . Table 11.3 is similar to Table 13.1 except that in Table 13.3 the error ratio is used since efficiently stratified ratio estimation is being discussed. The sample sizes shown in Table 13.3 are generally much smaller than in Table 13.1 because the error ratio is generally much smaller than the coefficient of variation for a given population.

Table 13.3: Required Sample Size Assuming a Large Population

		<i>Error Ratio er</i>						
		<i>0.4</i>	<i>0.5</i>	<i>0.6</i>	<i>0.7</i>	<i>0.8</i>	<i>0.9</i>	<i>1.0</i>
<i>Desired Relative Precision D</i>	<i>0.25</i>	7	11	16	21	28	35	43
	<i>0.20</i>	11	17	24	33	43	55	68
	<i>0.15</i>	19	30	43	59	77	97	120
	<i>0.10</i>	43	68	97	133	173	219	271

Assessing the Error Ratio without a Prior Sample

Table 13.3, above, illustrated four examples of relationships between x and y . These are typical examples of the type of association expected under the ratio model, assuming various error ratios. In each graph, the solid line represents the expected value of y given x , $\mu_i = \beta x_i$, and the broken lines represent the one standard deviation intervals around the mean: $y_i = \beta x_i \pm \sigma_i$. In preparing these graphs, the secondary equation $\sigma_i = \sigma_0 x_i^\gamma$ has been assumed with $\gamma = 0.8$.

In most impact evaluation studies, the error ratio can be expected to be in the range 0.4 to 1.0, as illustrated in Table 13.3. If the tracking system is expected to provide quite accurate estimates of the actual savings of most sample projects in the evaluation study, then the error ratio is likely to be relatively small, e.g., near 0.4. This might be the case, for example, if the program provides energy efficiency retrofits to large commercial buildings, and the tracking estimates of savings are based on a fairly detailed analysis of each project that is undertaken in the program. If the tracking system is expected to provide rather poor estimates of the actual savings of most sample projects in the evaluation study, then the error ratio is likely to be larger, e.g., near 1.0. This might be the case, for example, if the program is an express-style program that requires only a simple application and does not provide any site-specific analysis as part of the program delivery.

Estimating the Error Ratio from a Sample

When stratified ratio estimation is being used to analyze a sample, the sample can also be used to estimate the underlying error ratio for use in future sample designs. Assuming the secondary equation $\sigma_i = \sigma_0 x_i^\gamma$ with $\gamma = 0.8$, then the error ratio can be estimated as

$$\hat{e}r = \frac{\sqrt{\left(\sum_{i=1}^n w_i e_i^2 / x_i^\gamma\right) \left(\sum_{i=1}^n w_i x_i^\gamma\right)}}{\sum_{i=1}^n w_i y_i}.$$

Here, as usual, $e_i = y_i - b x_i$.³³⁸

Model-Based Stratification

The preceding results assume that stratified ratio estimation is used with an efficiently stratified sampling plan. This section will describe how to construct an efficiently stratified sampling plan. The goal is to group the projects into several strata based on the value of x , usually the tracking estimate of savings, and then specify the number of sample projects to be selected from each stratum. The following method is called model-based stratification by size.³³⁹

The following steps are required:

1. Create a spreadsheet or database listing each project in the population and providing the value of x_i for each project, $i = 1, \dots, N$.
2. Use the assumed secondary equation of the ratio model to calculate σ_i for each project, $i = 1, \dots, N$. Typically, $\sigma_i = \sigma_0 x_i^\gamma$ where γ is a set value, often 0.8. The value of σ_0 can be calculated from the assumed value of the error ratio using the equation given previously. Sort the list by increasing σ_i . For each $i = 1, \dots, N$ calculate the cumulative sum of the σ_i , $c_i = \sum_{j=1}^i \sigma_j$.
3. Choose the desired number of strata L , (usually three to five) and divide the projects in the sorted list so that the sum of the σ_i is approximately equal in each of the L strata. This can be done by calculating $h_i = \text{INT}\left(L \frac{c_i}{c_N} + 0.99999999\right)$.

³³⁸ If it is also necessary to estimate gamma from the sample, a method is available. See “Estimating regression models with multiplicative heteroscedasticity” (Harvey 1976).

³³⁹ Another method of constructing strata is called Dalenius-Hodges stratification by size. The authors have chosen to emphasize model-based stratification because it is known to provide nearly optimal sample designs for stratified ratio estimation. See *Model Assisted Survey Sampling*, (Sarndal et al. 1992).

Here the *INT* function rounds the value down to the nearest integer and 0.99999999 has been added to the equation to keep the last project from being assigned to a new stratum.

Once the strata have been constructed as just described, the sample should be allocated equally to each stratum. If the sample size in a particular stratum exceeds the population size in that stratum, the projects in that stratum should be selected with certainty. If desired, the sample may be increased in the remaining strata so that the sample size is closer to the planned value.

In some applications, it may be desirable to stratify the population by a categorical characteristic of the projects as well as by size. For example, the projects might be stratified by building type, technology, contractor, or region. The underlying principle is that the sample size allocated to each categorical stratum should be proportional to the sum of the σ_i within each stratum. Given the definition of the error ratio, a convenient way to determine the sum of the σ_i within each stratum is to multiply the expected actual savings in each stratum by the error ratio assumed in the stratum. This gives the rule: the sample size allocated to each categorical stratum should be proportional to the product of the expected actual savings in each stratum and the error ratio assumed in the stratum.³⁴⁰

Once the sample size has been determined within each categorical stratum, the projects within each stratum should be further stratified by size as described above.

The Expected Statistical Precision for Any Sample Design

This section discusses how to assess the expected statistical precision of the stratified ratio estimator when stratified ratio estimation is used with an arbitrary sample design. These results assume that the ratio model is accurate and that the sample design is truly followed without non-response or other similar problems.

To develop the result of interest, a new concept is needed. For any given sample design, define the inclusion probability π_i to be the probability that project i is included in the sample, for all $i = 1, \dots, N$ in the population. Assume that the inclusion probability is greater than zero for every project in the population, and that sample size n is fixed. There are two useful facts about inclusion probabilities. First, the population sum of the inclusion probabilities is equal to n . Second, for any stratified sample design, the inclusion probability is equal to the sampling fraction in each stratum.

Now the result: Let b be the stratified ratio estimator. Under the ratio model, the expected value of the standard deviation of b in repeated sampling is approximately

³⁴⁰ This result can be used to allocate evaluation resources among a portfolio of programs, especially if the marginal evaluation cost per sample project is approximately the same for all projects in the portfolio. See the chapter on Uncertainty.

$$sd(b) = \frac{1}{X} \sqrt{\sum_{i=1}^N (\pi_i^{-1} - c) \sigma_i^2}.$$

Here c is 1 if the finite population correction is desired, or 0 if not.³⁴¹ Under the ratio model, the expected relative precision can be defined to be $rp = sd(b)/\beta$.

The preceding equation can be used to assess the expected relative precision for any stratified sample design under the ratio model. This methodology can be used, for example, to explore the effect of increasing the number of strata. This type of analysis indicates that three to five model-based strata are adequate in most impact evaluation applications. This equation has also been used to explore the effect of using model-based stratification with a set value of gamma that is smaller than the value assumed in the ratio model. In several evaluation applications, it has been shown that there is very little loss in expected statistical precision if the strata are constructed using a gamma of 0.5 when the value in the secondary equation is 0.8. This tends to decrease the sampling fractions in the strata containing larger projects. This can sometimes facilitate recruiting and data collection.

Using the preceding equation, a sample design is said to be optimal under the assumed ratio model if the inclusion probabilities minimize $sd(b)$ for a given sample size n . It

can be shown that a sample design is optimal if and only if $\pi_i = n \sigma_i / \sum_{i=1}^N \sigma_i$ provided

this is not greater than 1. If $n \sigma_i / \sum_{i=1}^N \sigma_i > 1$, then project i should be selected with certainty.³⁴²

Applicability to Impact Evaluation

Stratified ratio estimate also relies on the assumptions that the sample design is followed and that the true savings are measured for the sample projects with little or no bias, as discussed in the section on simple random sampling. Since the sample can generally be smaller with stratified ratio estimation than with simple random sampling, it should be possible to give even more attention to minimizing bias from self-selection, non-response, deliberate substitution of sample projects, or systematic measurement error.

³⁴¹ For example, the finite population correction might not be suitable if random measurement error is a large contributor to the residual standard deviation of each project.

³⁴² Under the ratio model, $\sum_{i=1}^N \mu_i = \beta X$. This result can be used to show that if $\pi_i = n \sigma_i / \sum_{i=1}^N \sigma_i$

for all projects in the population and $c = 0$, then $rp = 1.645 \frac{sd(b)}{\beta} = 1.645 \frac{er}{\sqrt{n}}$. This justifies

our use of the error ratio to calculate the estimated relative precision assuming that a ratio estimator is used with an efficiently stratified sample design and a large population.

Stratified ratio estimation is generally especially effective when simple random sampling is inappropriate. Whenever the coefficient of variation of savings is greater than one, stratified ratio estimation should be considered. Stratified ratio estimation will almost always be more effective than simple random sampling if the program provides good tracking estimates of savings.

Stratified ratio estimation often focuses on the relationship between the tracking estimates of savings and the actual savings. The two key parameters are the realization rate and error ratio. The realization rate is the slope of the trend line. It is the ratio between the average or total value of the actual savings and the average or total value of the tracking estimates. Thus, the realization rate reflects the amount of systematic bias in the tracking estimates of savings.

The error ratio, on the other hand, describes the strength of the association between the tracking estimates of savings and the actual savings, i.e., the variation of actual savings around the trend line associated with the realization rate. The error ratio measures whether the tracking savings are accurate from project to project across the population of projects.

The error ratio is a useful indicator of the quality of the program delivery system. Well-designed and managed programs will tend to have smaller error ratios than programs with poorer control and less attention to detail. Indeed, if the error ratio is found to be higher than expected, it generally indicates that there is a problem with program delivery. Conversely, stratified ratio estimation tends to reward strong programs, i.e., those with relatively small error ratios, by making it possible to carry out an effective impact evaluation using a relatively small sample.

With stratified ratio estimation, the ratio model has been used to assist in the development of a suitable sample design. It is important to understand, however, that the model is only used to develop the sample design. The model is not used to support the statistical analysis of the sample data, except the estimation of the error ratio. If the model is accurate, the achieved statistical precision will be close to the expected statistical precision predicted by the model. If the model is inaccurate, the expected statistical precision may be inaccurate also. But even if the model is inaccurate, the stratified ratio estimator is still free of any material bias and the standard error is still a good guide to the achieved statistical precision.³⁴³

³⁴³ Sarndal has referred to these methods as model-assisted since, although the analysis does not depend on the accuracy of the model, the model does guide the analysis. (Sarndal et al. 1992), pp. 227 and 239. Sarndal provides a much more general model called the generalized regression model which may, in some circumstances, suggest other estimators such as the difference or regression estimators, but the authors have found that the ratio estimator generally is suitable in evaluation.

A Sampling Roadmap

Building on the preceding foundations, a sampling roadmap is now provided, i.e. one recommended way of approaching statistical sample design and analysis in a typical evaluation study. This map can provide effective guidance for most evaluation studies, but occasionally a special circumstance may arise that will call for some other approach.

Figure 13.5 provides an overview of the overall roadmap. Each step will be discussed below in sequence, starting with the population database.

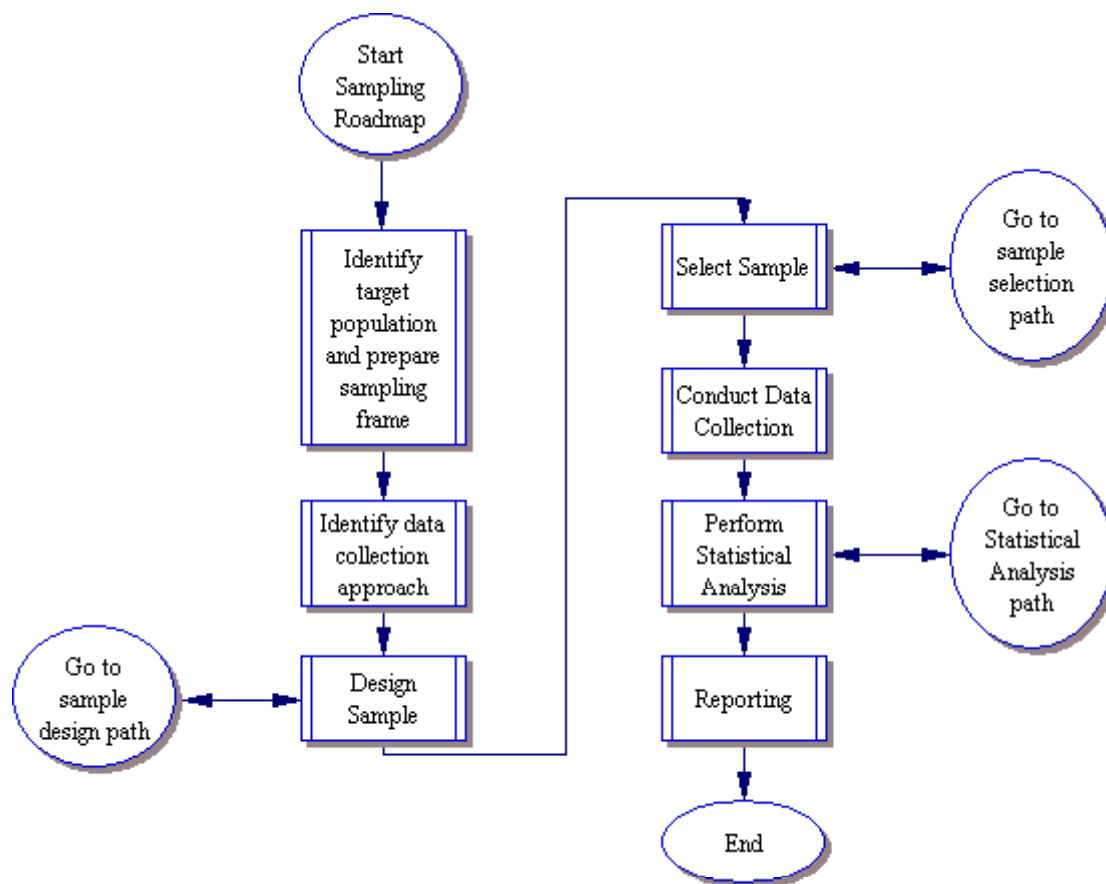


Figure 13.5: Overall Sampling Roadmap

The Population Database

The first step in sample design is usually to develop a population database (sometimes called the sampling frame) listing each unit in the population and providing relevant information for each unit. For example, in an impact evaluation study, the population may be the set of projects implemented in the program in a given year. In a baseline study the population may be the set of all non-residential customers served by a given utility or the set of all vendors of a given appliance in a specified region.

The population database can often be extracted from a program's tracking system³⁴⁴ or a utility's billing system. The sampling frame can also come from directories of businesses or other lists, as long as these provide a complete representation of the population being studied. The data should be imported into a spreadsheet such as Microsoft Excel or Lotus 123.³⁴⁵

Some consolidation or reorganization may be required to make the population database more suitable for sampling, data collection and M&V analysis. The population database should have a single record for each unit – project, customer, vendor, etc., for data collection and analysis. For each unit, the population database should include a unique ID, relevant descriptive information, and available contact information.³⁴⁶ The population database should also contain a suitable measure of the size of each unit such as the tracking estimate of the kWh savings of the project, the annual consumption of the customer, or the total sales of the vendor.

In addition to the population database itself, there usually is more detailed information for each unit in the population. For example, if the population is a set of projects, the project files should provide descriptive information for each installed measure and its physical location in enough detail that the measure can be located at the site during an on-site audit. This information does not have to be stored electronically but it should be available to the evaluation team for the projects that are included in the sample.

Step 1: Identify the target population and prepare the sampling frame.

Prepare a spreadsheet database listing all units in the population (projects, customers, etc.) with supporting information for each unit.

Each row should be a separate unit in the population. Separate columns should provide:

- ID
- Descriptive information
- Utility account number (to link the project with utility billing data)
- Measure of size (e.g., tracking estimate of savings or annual consumption from billing data)
- Contact information
- Other relevant information

³⁴⁴ An adequate program tracking system is essential to effective program delivery and control, in addition to effective evaluation. But the design of a tracking system is beyond the scope of this chapter.

³⁴⁵ While Excel will be used to illustrate the methods of data management and analysis discussed in this chapter, it is often advantageous to implement the analysis in SAS or another suitable statistical analysis system.

³⁴⁶ The completeness and accuracy of the contact information will affect the cost of data collection. Depending on the type of program and type of evaluation study, you may need, for example, contact information for the facilities' manager or O&M supervisor, or suitable e-mail addresses if an e-mail survey is planned.

Sample Design

Figure 13.6 summarizes the sample design path of the sampling roadmap. In impact evaluation, there are three primary sampling approaches:³⁴⁷

1. A census,
2. Simple random sampling, or
3. Stratified ratio estimation.

Census

A census of all units in the population may be used if the population is very small (i.e., 50 or fewer units) and the unit cost of data collection is so low that the added complexity of sampling is not justified. The use of a census does not mitigate the danger of bias from non-response, measurement error, etc. Therefore the decision to attempt a census should not be made lightly. It is often better to use sampling to limit the data collection to a smaller number of units so that added attention can be devoted to each unit.

Simple Random Sample

A simple random sample may be the most reasonable method if the coefficient of variation of the size of the units in the population is less than one and the unit cost of data collection is so moderate that the added complexity of stratified ratio estimation is not justified.

³⁴⁷ These approaches are expected to handle most applications. Other methods such as cluster sampling, multi-stage sampling, and sequential sampling may be advantageous in some circumstances.

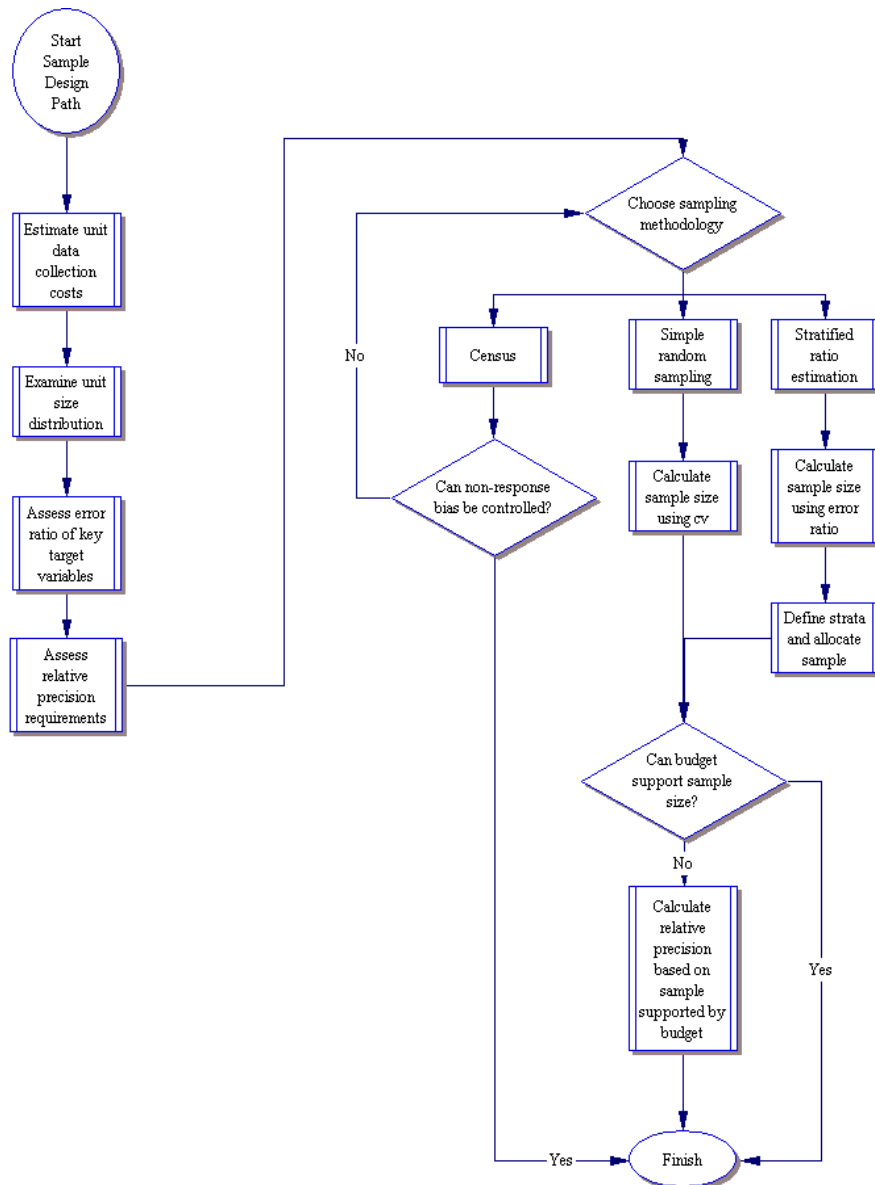


Figure 13.6: Sample Design Path

In this case the primary sample design decision is the choice of the sample size. For any assumed sample size, the expected statistical precision can be estimated from the coefficient of variation and the number of projects in the population. The sample size can be chosen by considering the unit cost of data collection, the available budget, and the expected statistical precision. If the desired statistical precision is known, the sample size can be calculated directly.

Stratified Ratio Estimation

Stratified ratio estimation will usually provide improved statistical precision or reduced sample sizes compared to simple random sampling, provided that a measure of variability called the error ratio is smaller than the coefficient of variation. In this case, stratified

ratio estimation may be the preferable method unless the reduction in sample size does not warrant the added complexity. The error ratio should be taken from a prior study of the program, from a similar study, or from an assessment of the accuracy of the tracking estimates.

For any assumed sample size, the expected statistical precision can be calculated from the error ratio and the number of projects in the population. As under simple random sampling, the sample size can be chosen by considering the unit cost of data collection, the available budget, and the expected statistical precision. If the desired statistical precision is known, the sample size can be calculated directly.

With stratified ratio estimation, it is necessary to construct strata reflecting the size of each unit in the population, e.g., the tracking estimate of savings or the annual consumption of the customer. If desired, qualitative strata can also be developed.

Steps in Developing the Sample Design

The following discussion describes the sample design steps shown in Figure 13.6.

Step 1: Develop the general approach to the data collection and project-specific M&V analysis.

- | |
|--|
| <ul style="list-style-type: none">a) Develop the most appropriate approach to the data collection and M&V analysis, e.g., telephone survey, on-site audit, site-specific simulation, etc.b) Estimate the average cost for the typical unit in the population. You may need to estimate the sample size as specified below to assess the average cost.c) If there are less than 50 units in the target population and the data collection costs are moderate, a census may be planned and the following steps may be omitted. |
|--|

Step 2: Examine the distribution of size of the units in the population.

- | |
|---|
| <ul style="list-style-type: none">a) Select a suitable measure of the size of each unit. If the population is a set of projects, the tracking estimate of savings should be used. For non-participant surveys and billing analysis, annual consumption may be used.b) Calculate the mean, standard deviation and coefficient of variation of size. The coefficient of variation is the standard deviation divided by the mean.c) If the coefficient of variation is less than 1 and the data collections costs are moderate, then simple random sampling can be used and Step 3 and Decision 1 can be omitted.d) This step may be omitted if stratified ratio estimation is planned. |
|---|

Step 3: Assess the error ratio for the key target variable(s) to be collected in the study.

- a) If possible, use an error ratio from a similar prior study of this program or a similar program.
- b) If no prior experience, assume an error ratio of 1.
- c) In most applications, the tracking kWh savings should be used as the x-variable.

Decision 1: Choose the type of sampling methodology, either simple random sampling or stratified ratio estimation.

Criteria:

- If the error ratio from Step 3 is smaller than the coefficient of variation from Step 2, use stratified ratio estimation unless the reduction in sample size does not offset the added complexity of sample design and analysis.
- Otherwise, use simple random sampling.
- If using simple random sampling, complete steps 4 and 5 and then go on to sample selection.

Step 4: Assess the desired relative precision.

Criteria:

- The size of the program,
- The error ratio or coefficient of variation in the program,
- The total evaluation resources available within the portfolio,
- The optimal allocation of evaluation resources among the programs in the portfolio.

Notes:

- a) It can be helpful to calculate the expected relative precision for various sample sizes.
- b) If simple random sampling and the 90% level of confidence are used, the expected relative precision is given by the equation $rp = 1.645 \frac{cv}{\sqrt{n}}$

provided that the number of units in the population, denoted N , is large compared to the planned sample size, n . Otherwise

$$rp = 1.645 \sqrt{1 - \frac{n}{N}} \frac{cv}{\sqrt{n}}. \text{ Here } cv \text{ is the population coefficient of variation}$$

of the target variable of interest, which generally can be estimated as the coefficient of variation of the measure of size in the population.

- c) If stratified ratio estimation is used, the preceding equation can be used with the error ratio instead of the coefficient of variation.
- d) This step may be skipped if the evaluation budget is already given for the study being planned.

Step 5: Calculate the required sample size to achieve the desired relative precision.

- a) If the evaluation budget is already given for the study being planned, the sample size can be calculated from budgetary considerations. In this case, the expected relative precision should be calculated and documented.
- b) If the desired relative precision has been specified, the sample size should be calculated. If simple random sampling is used and if D is the desired relative precision at the 90% level of confidence, then the required sample size is $n = \left(\frac{1.645 cv}{D} \right)^2$ provided that the number of units in the population, denoted N , is large compared to the planned sample size, n . Otherwise, calculate n_0 using the preceding equation, and then apply the finite population correction $n = \frac{n_0}{1 + n_0/N}$.
- c) If stratified ratio estimation is used, then the preceding equations can be used with the error ratio instead of the coefficient of variation.
- d) If the indicated sample size is not sufficiently smaller than the population to justify the added complexity of sampling, a census may be used. The use of a census does not mitigate the danger of bias from non-response, measurement error, etc.

Decision 2: If stratified ratio estimation is used, will qualitative strata be used?
Yes or No

The target population can, if desired, be divided into two or more qualitative strata based on any information that is available for all units in the population, e.g., building type, type of measures installed, climate zone.

Criteria:

- Qualitative stratification may provide added control over the distribution of the sample.
- Qualitative stratification may improve statistical precision if error ratios vary substantially between qualitative strata.
- Sample size in each qualitative stratum should not be too small (e.g., less than 25), so usually only possible if a relatively large sample is planned.
- Qualitative stratification usually adds complexity to the analysis.
- In the analysis, each qualitative stratum can be considered individually and the results combined across the strata using methods similar to the methods discussed for assessing the total across a set of projects within a portfolio.
- Meaningful results can often be developed for some subsets of the sample even if these subsets are not set up as qualitative strata in the sample design.

Step 6: If stratified ratio estimation, develop qualitative strata (optional).

- a) Divide the units in the population into unique qualitative strata based on any relevant characteristics (type of building, climate zone, etc.).
- b) Tabulate the number of units and total size (e.g., tracking savings) in each qualitative stratum.
- c) Assess the error ratio in each stratum (may be equal to the overall error ratio) and allocate the sample to each stratum in proportion to the product of the total size and the error ratio in each stratum.

Step 7: Construct the size strata (required under stratified ratio estimation).

- 1) Choose the number of size strata, denoted L , to be used overall or within each qualitative stratum (if used). Usually L is in the range 3-5.
- 2) Use the values of size in the population database to construct the strata overall or within each qualitative stratum (if used).
- 3) Construct the strata:
 - a) Assume a value for the beta parameter of the ratio model, usually $\beta = 1$ and calculate $\mu_i = \beta x_i$ for each case in the population database. Here x_i is the measure of size of case i .
 - b) Assume a value for the gamma parameter of the ratio model, usually $\gamma = 0.8$ and calculate x_i^γ for each case in the population database. Here x is the measure of size of each project i .
 - c) Calculate the σ_0 parameter of the ratio model using the equation

$$\sigma_0 = er \frac{\sum_{i=1}^N \mu_i}{\sum_{i=1}^N x_i^\gamma}. \text{ Here } er \text{ is the error ratio from step 3.}$$

- d) Calculate $\sigma_i = \sigma_0 x_i^\gamma$ for each case in the population database.
- e) Sort the population database by increasing values of σ_i .
- f) For each case i in increasing order, calculate the cumulative sigma

$$c_i = \sum_{j=1}^i \sigma_j.$$

- g) Divide the projects in the sorted database so that the sum of the σ_i is approximately equal in each of the L strata. In Excel this can be done by calculating $h_i = INT\left(L \frac{c_i}{c_{Ni}} + 0.99999999\right)$.

- 4) Allocate the overall sample or the sample allocated to each qualitative stratum (if used) equally among the size strata in each stratum.

Sample Selection

Figure 13.7 summarizes the sample selection path. As in the preceding example, the population database is assumed to be in a spreadsheet such as Microsoft Excel or Lotus 123. Each row of the spreadsheet must be a separate sampling unit such as a project or customer. The sample can be selected by creating a new column with the value of a random number assigned to each row, sorting the rows according to the random variable, and designating the first rows in the sorted database to be the sample.

With Simple Random Sampling

If simple random sampling is used, the following steps can be followed to select the units to be included in the sample and to prioritize the units to be used if any replacements are required.

Step 1: Assign a random number to each unit in the population database.

- Each row in the spreadsheet must be an individual sampling unit.
- In Excel, the function RAND() can be used to assign a random number to each row, using a new column.
- In Excel, after the random numbers have been created, their values must be frozen using edit - copy and edit - paste special - values.

Step 2: Sort the units in the sampling frame according to the increasing values of the random number created in Step 1.

Step 3: Use a new column to assign the priority to each unit in the sampling frame in the order established in Step 2. The first unit should have priority 1, the second unit priority 2, etc.

Step 4: Designate the units to be included in the primary sample, and designate the backup units.

- Let n denote the number of units to be included in the sample. The primary sample is the first n units in the sampling frame in the order established in Step 2. In other words the sample is the units with priority from 1 to n .
- The replacement or backup units are the next n units, to be used in the order established in Step 2. In other words, the unit with priority $n+1$ is the first replacement unit, the unit with priority $n+2$ is the second replacement unit, etc.
- The primary sample should be used initially. A replacement should only be used when a primary sampling unit cannot be used. Replacements should be minimized to control the risk of selection bias. All replacements should be thoroughly documented.

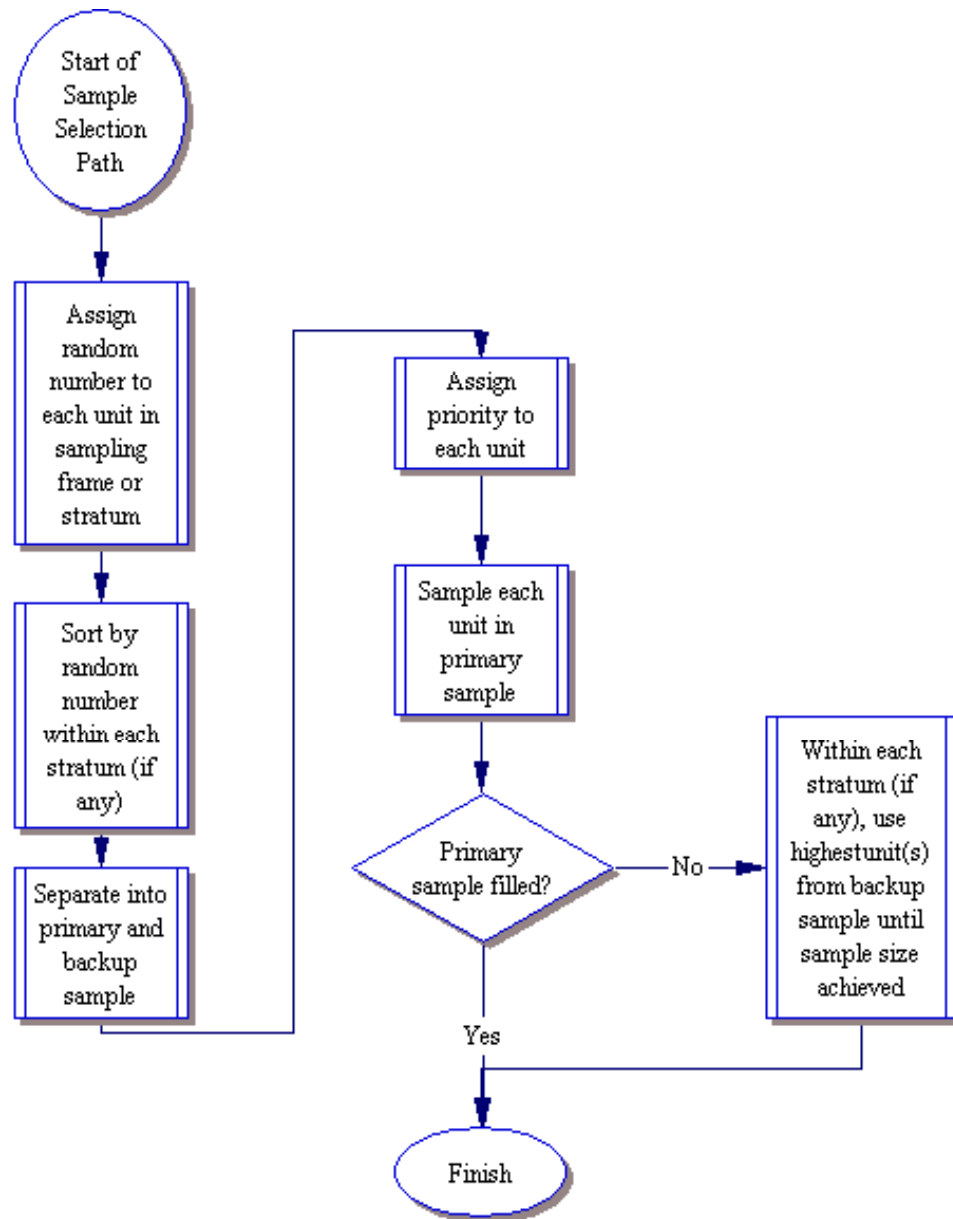


Figure 13.7: Sample Selection Path

With Stratified Ratio Estimation

If stratified ratio estimation is used, the preceding steps can be carried out within each stratum. In particular, the population database should be sorted by stratum and then by the random number within each stratum. In each stratum, the first unit should have priority 1, the second unit priority 2, etc. To the extent possible replacements for any substitutions should be made from within the same strata. If no replacements are available within a given size stratum, the replacement should be taken from the immediately preceding size stratum.

Data Collection and M&V

Once the sample has been selected, the focus switches to the data collection and M&V analysis for each sample unit. The work will vary greatly from one study to another. Some of the data collection and analysis methods include file reviews, telephone surveys, free ridership data collection and analysis, on-site audits, billing data analysis, computer simulation, and short-term monitoring. The most suitable approach will vary depending on the type of evaluation being conducted (impact, process, market) and from study to study.

As discussed in the Uncertainty chapter of this Framework, most measures of statistical precision assume that the data collection and M&V work is essentially free of both non-response and measurement bias. Therefore it is essential to minimize bias and to report any potential risks of bias in both the recruiting and site-specific data collection and analysis.

Recruiting and Scheduling

If a study involves on-site data collection, the recruiting and scheduling of the sample sites is usually a separate task. This is usually done over the telephone. A special database is generally used to track each attempt to contact the customer. The recruiting staff should be skilled and well-trained in:

- Identifying the person(s) that can provide permission for the work and arrange access to the site,
- Explaining the purpose of the study, the information that is required, and the data collection procedures that will be followed,
- Collecting any added information that may be required such as the location of the measures, the best time for the on-site visit, and the contact person for the visit, and
- Understanding the issue of non-response bias, and how to use the backup sample appropriately.

Using Backups

It is very important to use the backup sample correctly. The most efficient way to recruit a sample of the desired size may appear to be to contact both the primary and backup sample at once and to schedule those sites that are first to respond and agree. But this is generally not sound practice since this approach ensures that the response will be no better than 50%, assuming that the backup sample size is equal to the primary sample size. Instead, the initial recruiting effort should be limited to the primary sample. A backup should be used only if a primary sample site is impossible to contact or refuses to participate.³⁴⁸

³⁴⁸ Sometimes the primary sample site is found to be out of scope. This situation is beyond the scope of this discussion.

The Sample Database

The statistical analysis can often be carried out in a spreadsheet such as Excel or 123. A sample database is created with one row for each unit in the sample and with a separate column for each field of collected data or M&V results.³⁴⁹ In a typical impact evaluation study, the sample database may include the project ID, the tracking estimate of savings, the gross savings from the M&V analysis, and the net savings from the free ridership analysis.

Quality Control

The goal of quality control is, to the extent possible, to ensure that the sample is free of selection bias and that the collected data are sound and reliable. Three techniques for quality control are especially related to the statistical analysis:

- Monitoring the recruiting, scheduling, and use of backups,
- Reviewing the data that is collected and the M&V analysis that is done for each sample case, and
- Identifying the most influential sample cases.

In order to ensure that problems are detected early enough for timely correction, each of these techniques generally should be done during pre-testing, again soon after data collection has begun, and repeated periodically through the data collection stage of the study.

The primary purpose of monitoring the recruiting, scheduling, and use of backups is to avoid bias from non-response or substitutions and to ensure that the conventional measures of statistical precision are not misleading. As discussed elsewhere, the standard statistical analysis procedures rest on the assumption that the sample is selected strictly following the sample design. Therefore it is important to monitor any deviations from the sample design, e.g., inability to contact the customer due to inaccurate contact information, refusal of the customer to participate in the study, refusal or inability to provide all of the information required for the analysis, etc. If problems are observed, the procedures for implementing the study may need to be adjusted. For example, an incentive may be needed for participating in the study, the recruiting staff may need to be given added training, or the survey may need to be simplified.

The second type of quality control work involves reviewing the data and M&V analysis for each sample case. For example, inconsistencies in the responses to various survey questions may indicate problems with the wording of the question. Similarly, estimates of savings developed in the M&V analysis should be compared to billing data. These issues are discussed further in the M&V section.

It is generally useful to identify the sample cases that have the greatest influence on the results. A close examination of these cases may shed insights into the program itself as

³⁴⁹ Often, a separate worksheet is used to do the M&V analysis for each sample site and the key results are linked to the sample database.

well as into the data collection and M&V analysis procedures. Two specific techniques can be mentioned – scatter plots and weighted squared errors.

Scatter plots and other graphical techniques can be used to examine the key data elements across the sample. In an impact evaluation, for example, it is generally useful to examine a scatter plot relating the measured savings to the tracking savings. This type of scatter plot can display patterns in the data, particular sample cases that are outliers, and, sometimes, subsets of cases that deviate systematically from the general pattern. These scatter plots can also shed light on the validity of the ratio model that is the foundation for stratified ratio estimation.

With a standard scatter plot, it can be difficult to factor in the case weights that play a role in stratified ratio estimation. The weighted squared error technique addresses this issue. This technique builds on the methods of stratified ratio estimation that are discussed in the following section, but it can be briefly described here. The idea is to calculate the weighted squared error that reflects the contribution of each sample case to the standard error of the ratio estimator, i.e., $w_i(w_i - 1)e_i^2$. The preceding notation is defined in the next section. The sample cases with the largest weighted squared error can be regarded to be the most influential cases. By sorting the sample data according to the value of the weighted squared error, these cases can be easily identified and examined more closely. This work can easily be carried out in a spreadsheet and summarized either in a table or using a suitable graphical technique.

Statistical Analysis

The appropriate statistical analysis will vary depending on the goals and special characteristics of each study. This section will emphasize the methods used to extrapolate the results from the sample to the population and to estimate an error bound for the results. If stratified ratio estimation is used, it is also important to use the sample data to estimate the error ratio to help plan future studies. Figure 13.8 summarizes the steps in the statistical analysis path.

Regardless of the sample design path – census, random, or ratio – the issue of bias should be addressed in the analysis. If a census was attempted and has been essentially complete, or if sample data have been developed for essentially all of the original sample units with almost no refusals or replacements, and if there is no material measurement error in data collection and site-specific analysis, then the statistical analysis is relatively straightforward. Otherwise, the analysis and report should address the possibility of bias.

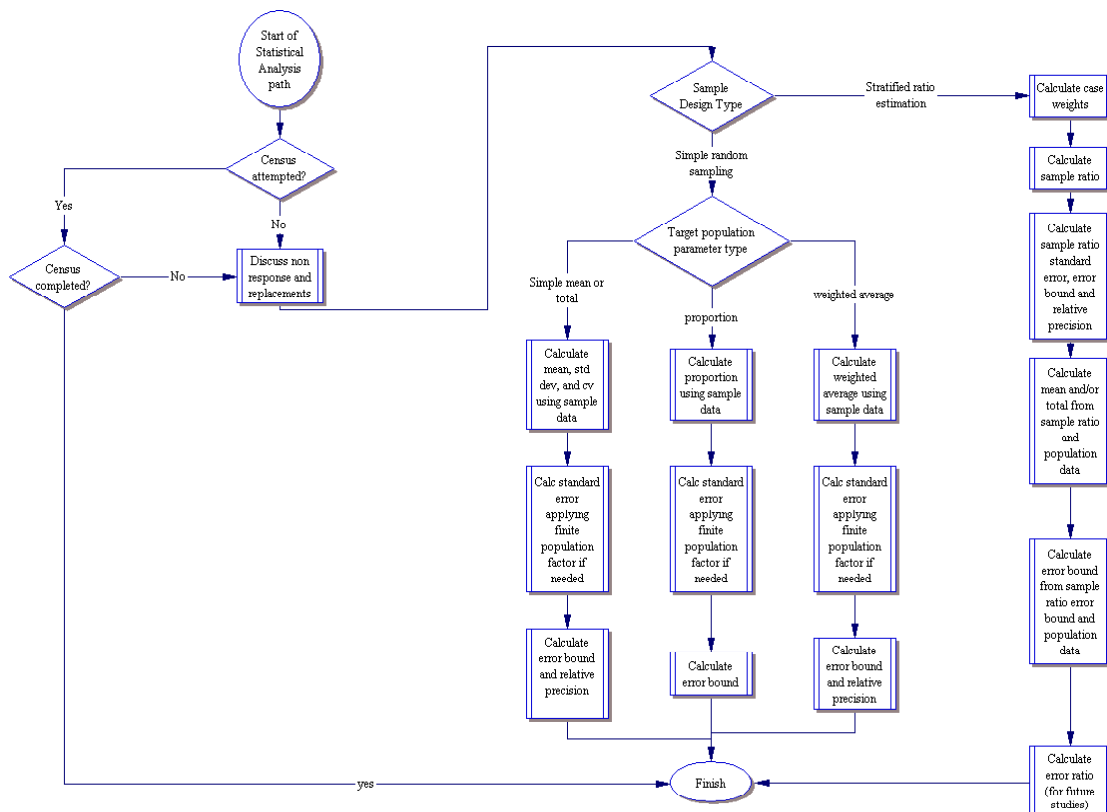


Figure 13.8: Statistical Analysis Path

With a Census

If a census has been taken of all units in the population, the results can be reported with no special statistical analysis. For example, the total evaluated savings and average savings per project can be calculated and reported from the sample data.

If a census was attempted but was not completed due to refusals or other reasons, then the analysis is more complex. The key decision is whether the included units can be regarded as representative of the units that were omitted from the census.

For example, in a persistence study, if a census was attempted but some program participants were excluded from the study because they could not be located, it may be more appropriate to write off the corresponding savings than to assume that these participants are similar to the participants included in the sample.

On the other hand, if the exclusion of the project from the census can be assumed to be statistically independent of the information to be collected, then the projects included in the study may be regarded as if they are a true sample. For example, suppose that in a persistence study, some program participants could not be included in the study because they refused to cooperate. If it seems plausible that the refusal was statistically independent of the persistence of the savings, then the participants that did cooperate can be regarded as a sample of all participants. In this case, either simple random sampling

or stratified ratio estimation should be used to analyze the data, depending on the coefficient of variation and error ratio.

With Simple Random Sampling

If simple random sampling has been used and there have been a limited number of replacements or non-respondents, the analysis is usually relatively straightforward. Otherwise, the disposition of the original sample should be discussed. Any non-respondents or replacements should be explained. The possibility of selection or non-response bias should be discussed in the report.

The methodology used in the statistical analysis depends on whether the parameter to be estimated is a mean or total, a proportion, or a weighted average or ratio. A mean or total is often used to summarize a quantitative variable such as the measured savings of each project in the population. A proportion is often used to summarize a qualitative variable such as whether or not a program participant is satisfied with a particular measure. A weighted average is often more appropriate than a simple mean, especially if the units vary substantially by size. For example the average efficiency of a population of air conditioners is generally defined to be the weighted average of the efficiency of each air conditioner, using a measure of the size of each unit such as tons as the weight. A ratio is often of interest in evaluation. For example, the gross realization rate is the ratio between the total gross savings of all projects in the population and the total tracking savings.

A population mean can be estimated by calculating the associated sample mean. The corresponding standard deviation and coefficient of variation should also be calculated. Then the standard error, error bound and relative precision are calculated in order to characterize the statistical precision of the sample mean as an estimator of the population mean. A population total can be estimated by multiplying both the sample mean and the corresponding error bound by the number of units in the population.

A population proportion can be estimated by calculating the associated sample proportion. Again the standard error and error bound are calculated in order to characterize the statistical precision of the sample proportion as an estimator of the population proportion.

A weighted average can be estimated using the corresponding weighted average of the sample data. As usual, the standard error, error bound and relative precision are calculated in order to characterize the statistical precision of the sample weighted average as an estimator of the population weighted average. A ratio can also be estimated using these same techniques.

Decision 1: Is the target population parameter a simple mean or total, a proportion or a weighted average? Go to path 1, 2 or 3 respectively.

Path 1: For a simple mean or total:

- Calculate the mean, sample standard deviation and coefficient of variation using the sample data.
- Calculate the standard error using the equation $se = \frac{s}{\sqrt{n}}$. Here s is the sample standard deviation and n is the number of units in the sample.
- If the sample is more than 5% of the population, multiply the preceding result by the finite population correction factor $\sqrt{1 - \frac{n}{N}}$. Here N is the number of units in the population.
- Calculate the error bound at the 90% level of confidence using the equation $1.645 se$.
- Calculate the relative precision by dividing the error bound by the mean.
- If the population total is to be estimated, multiply the sample mean and the error bound by N , the number of units in the population.

Path 2: For a proportion:

- Calculate the proportion, denoted $\hat{p}$, using the sample data.
- Calculate the standard error using the equation $se = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$.
- If the sample is less than 5% of the population, multiply the preceding result by the finite population correction factor $\sqrt{1 - \frac{n}{N}}$. Here N is the number of units in the population.
- Calculate the error bound at the 90% level of confidence using the equation $1.645 se$.

Path 3: For a weighted average or ratio:

- Write the weighted average or ratio in the population as $B = \frac{\sum_{i=1}^N y_i}{\sum_{i=1}^N x_i}$. For example, if the ton-weighted average efficiency of a population of air conditioners is of interest, x_i is the size in tons of air conditioner i and y_i is the product of x_i and the efficiency of air conditioner i .
- Calculate the corresponding sample weighted average or ratio using the equation $b = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i}$.
- Calculate the standard error as $se = \frac{\sqrt{\frac{\sum_{i=1}^n (y_i - b x_i)^2}{n(n-1)}}}{\frac{1}{n} \sum_{i=1}^n x_i}$.
- If the sample is more than 5% of the population, multiply the preceding result by the finite population correction factor $\sqrt{1 - \frac{n}{N}}$. Here N is the number of units in the population.
- Calculate the error bound at the 90% level of confidence using the equation $1.645 se$.
- Calculate the relative precision by dividing the error bound by the sample weighted average.

With Stratified Ratio Estimation

Within each stratum, the final sample should be compared to the planned sample and any non-respondents or replacements should be reported and explained. Then the case weight should be calculated in each stratum and assigned to each of the sample projects in the stratum.

Just as with simple random sampling, the stratified ratio estimation path can be used to estimate means and totals, proportions, and weighted averages or ratios. The starting point is the estimation of a population ratio by calculating the case-weighted sample ratio. The standard error, error bound and relative precision are calculated in order to characterize the statistical precision of the sample ratio as an estimator of the population ratio.

Finally, the sample data should be used to estimate the error ratio – the key measure of the strength of the association between x and y in the population for planning future studies using stratified ratio estimation.

As an example, the preceding methods can be used to estimate the gross realization rate which is defined to be the ratio between the true gross savings (y) of all projects in the population and the total tracking savings (x). The error bound can be calculated to summarize the statistical precision of the sample realization rate, and the error ratio can be calculated to help plan future evaluation studies.

Once the results have been developed for the ratio, they can usually be used to estimate other parameters that may be of interest. For example, the population mean or total of the variable used in the numerator of the ratio (y) can be estimated by multiplying both the sample ratio and the corresponding error bound by the population mean or total of the variable used in the denominator of the ratio (x), provided that the latter is known. For example, the average gross savings or the total gross savings of all projects in the population can be estimated by multiplying the gross realization rate and the corresponding error bound by the average tracking savings or total tracking savings in the population.

The stratified ratio estimator is also the key to the estimation of results for a particular subset of the population. Suppose for example that an estimate of the gross realization rate is desired for the projects of a particular vendor or those that fall in a given region or market segment. This can be equated to a ratio in which the y -variable is the gross savings of each project in the subset of interest and zero otherwise, and the x -variable is defined to be equal to the tracking savings of each project in the subset of interest and zero otherwise. With these definitions, the usual stratified ratio estimation methods apply. In addition, the total gross savings of the projects of interest can be estimated by multiplying the estimated realization rate by the total tracking savings of the projects.

The preceding methods can also be used to estimate proportions. For example the proportion of all gross savings that is associated with a particular subset of the population can be estimated by dividing the total gross savings of the projects of interest, estimated as described in the preceding paragraph, by the total gross savings of all projects.

Step 1: Calculate the case weight of each sample project, denoted w_i .

- Tabulate the number of population projects in each stratum.
- Tabulate the number of sample projects in each stratum.
- Calculate the case weight as the number of population projects divided by the number of sample projects, and apply the results to each unit in the sample.

Step 2: For any population ratio or proportion, use the sample data and case weights to calculate the sample ratio.

- The population ratio or proportion is defined to be of the form $B = \frac{\sum_{i=1}^N y_i}{\sum_{i=1}^N x_i}$

where x and y are any suitable variables.

- The sample ratio is calculated as $b = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i x_i}$. Here w_i is the case weight calculated in step 1.

Step 3: Calculate the standard error of the sample ratio.

- For each unit i in the sample, calculate $e_i = y_i - b x_i$.
- Calculate the standard error of the sample ratio using the equation
$$se(b) = \frac{\sqrt{\sum_{i=1}^n w_i (w_i - 1) e_i^2}}{\sum_{i=1}^n w_i x_i}.$$
- Calculate the error bound at the 90% level of confidence using the equation $1.645 se$.
- Calculate the relative precision by dividing the error bound by the sample ratio.

Step 4: Calculate the corresponding error ratio for use in future sample designs.

- Assume a value for the gamma parameter of the ratio model, usually $\gamma = 0.8$
- For each unit i in the sample, calculate $e_i = y_i - b x_i$ as in step 2.
- For each unit i in the sample, calculate x_i^γ .
- Estimate the error ratio between x and y in the population using the

$$\text{equation } \hat{er} = \frac{\sqrt{\left(\sum_{i=1}^n w_i e_i^2 / x_i^\gamma \right) \left(\sum_{i=1}^n w_i x_i^\gamma \right)}}{\sum_{i=1}^n w_i y_i}.$$

Step 5: If the population mean or total of y is of interest, and the population mean or total of x is known, use the sample ratio to obtain the corresponding estimate and error bound.

- Estimate the population mean or total of y by multiplying the population mean or total of x by the sample ratio developed in Step 1.
- Calculate the corresponding error bound by multiplying the population mean or total of x by the error bound developed in Step 2.

Step 6: Sometimes there is an interest in the population ratio for a particular

subset S of the population, $B^* = \frac{\sum_{i \in S} y_i}{\sum_{i \in S} x_i}$. Here $i \in S$ is read, “ i is an element of S ”.

- Define an indicator variable I_i to be 1 if i is an element of S , and 0 otherwise.
- Define $y_i^* = I_i y_i$ and $x_i^* = I_i x_i$.
- Redefine the population ratio of interest to be $B^* = \frac{\sum_{i=1}^N y_i^*}{\sum_{i=1}^N x_i^*}$.
- Estimate B^* as specified in steps 2 and 3.
- Estimate the population total of y for the subset S as specified in step 5.
- Then, if desired, calculate the proportion of the grand total of y that is associated with the subset S .

Reporting Guidelines

There should be some standardization of how results are reported. The present discussion recommends reporting guidelines from the point of view of sampling and statistical analysis. The following guidelines are general and should be tailored to each project.

Findings

This section of the report will generally start by discussing the objectives of the study and characterizing the population using information taken from the population database. For example, in the case of an impact evaluation study, the number of projects and tracking estimates of savings should be described. Then the findings should be presented, with confidence intervals provided for the important results at the 90% level of confidence. Finally, the relevant sample-design parameters such as the coefficients of variation and/or error ratio should be reported for the key results to guide the planning of future studies.

Data Collection and M&V Methodology

This section should describe all data collection and M&V methodology used in the study. All data collection instruments should be provided, usually in an appendix to the report. An assessment of the risk of measurement bias should be included in this section.

Statistical Methodology

This section should document the sampling path: e.g., census, random, or ratio; give the parameters used to develop the sample design such as the assumed coefficient of variation and/or error ratio; and discuss the expected relative precision at the 90% level of confidence. This section should also discuss any substantial difference between the assumed sample design parameters and the estimates found from the sample data for the key results. If a stratified sample design is used, each stratum should be described, giving the population statistics for each stratum (number and total size), and the specified number of sample projects in each stratum.

The disposition of the sample should also be discussed, documenting all non-responses, refusals, substitutions, etc. If a stratified sample design was followed, this information should be provided for each stratum. If a census or random sample has been attempted, the report should describe any deviations from the intended sample by size category to help shed light on any selection bias. In any case, the report should discuss the underlying reasons that are believed to have contributed to the incomplete execution of the intended sample selection. Building on this information, the danger of selection bias should be discussed.

Finally the statistical analysis methodology should be described. Any deviations from the methodology in this chapter should be discussed.

Appendix to Chapter 13

Example and Use of Stratified Ratio Estimation for Sampling

This appendix provides an example of the stratified ratio estimating procedure, following the step-by-step sampling roadmap. The description of the methodology and steps are not repeated here in the interest of brevity. This example can be followed alongside the material above.

In this particular program, lighting retrofits were implemented in 1,248 small commercial sites. Under the program, efficient lighting was installed in relatively small non-residential sites. Very few lighting control measures were installed.

The objective of the study was to select a sample of projects from the program, assess the gross and net savings of each sample project, and then extrapolate the findings back to the population of all 1,248 projects in the population.

A population database was developed from the program tracking system. Table 13.4 shows a small portion of the population database. The table shows the first ten projects in the database as well as the last nine projects. For confidentiality reasons, the customer account number and contact information have been deleted from the database.³⁵⁰

Step 1 – Plan the General Approach to Data Collection and M&V

Based on prior experience, it was felt that on-site auditing and M&V engineering analysis would be the best approach to evaluating the energy and demand savings of this particular program. In a prior study of this program, short-term monitoring was used quite extensively.³⁵¹ In the prior study the hours of use estimated from the monitoring were very highly correlated with reported hours of use developed on a space-specific basis with almost no bias. So in the current project, no end use monitoring was planned.

However, fairly detailed on-site data collection and project-specific M&V analysis were planned. For each project in the sample, the analysis would start with a detailed review of the project file and review of the methods used to calculate the savings. Any apparent errors would be corrected and the savings recalculated. Then each sample site was to be visited, and a detailed on-site audit would be carried out with a focus on the installed lighting equipment. The type of fixtures and the associated wattage would be recorded and compared to the information in the file. The savings would be recalculated if any differences were observed. Then the number of each fixture type that was actually installed would be compared to the reported quantity and the savings would be

³⁵⁰ The authors are grateful to Lisa Shea and NSTAR Electric & Gas for allowing the use of this example. Minor modifications have been made to simplify the example. But the example accurately reflects an actual study carried out for NSTAR by RLW Analytics.

³⁵¹ “Evaluating the Underserved Small C&I Market: Building a Bridge to Implementation,” (Ledyard 2003).

recalculated. Finally the hours of use of each type of space served by the installed lighting would be assessed and the kWh savings would be recalculated. The result was taken to be the gross savings.

Table 13.4: The Population Database

Project	Vendor	Trk kWh
1	4	3,117
2	4	3,056
3	2	2,668
4	7	14,710
5	4	11,142
6	4	1,482
7	7	3,974
8	2	49,782
9	2	45,659
10	4	4,206
11	4	1,750
12	2	6,369
1238	4	51,579
1239	2	28,496
1240	4	11,737
1241	4	2,000
1242	4	129,205
1243	4	16,520
1244	4	44,611
1245	4	106,948
1246	4	624
1247	4	18,188
1248	4	52,225

Based on the file review, the on-site visit and an interview with the program participant, the free ridership would be assessed for each measure installed in the project. This information would be applied to the gross savings to calculate the net savings adjusted for free ridership. Finally, all of these sources would be used to identify any efficiency measures taken that were not rebated through the program but were the result of the program, and the savings would be recalculated adjusted for the participant spillover.

Similar information would be developed for the summer and winter (as applicable) demand savings, but these results will not be discussed here due to space limitations.

It was estimated that the preceding M&V work would cost about \$1,000 per sample project or about \$50,000 if the sample contained 50 projects.

Step 2 – Examine the Distribution of Size of the Projects in the Population

The kWh savings recorded in the tracking system was taken to be the measure of the size of each project. This variable is referred to as the tracking savings. In the database it is labeled as Trk kWh.

Table 13.5 summarizes the tracking savings provided in the population database. The table shows that the 1,248 projects had a total tracking savings of 20,119,315 kWh. They have an average tracking savings of 16,121 kWh per project and a standard deviation of 23,333 kWh per project. The coefficient of variation of the tracking savings is 1.45. The smallest project has 48 kWh of tracking savings and the largest project has 345,350 kWh of tracking savings.

Table 13.5: Summary of Tracking Savings in Population Database

Population Size	1,248 Projects
Total Tracking Savings	20,119,315 kWh
Average Tracking Savings	16,121 kWh per Project
Standard Deviation	23,333 kWh per Project
Coefficient of Variation	1.45
Minimum Tracking Savings	48 kWh
Maximum Tracking Savings	345,350 kWh

Table 13.6: The Number of Projects by Tracking Savings

Count of Project	
Trk kWh	Total
1-10000	691
10001-20000	269
20001-30000	100
30001-40000	65
40001-50000	38
50001-60000	26
60001-70000	19
70001-80000	7
80001-90000	4
90001-100000	8
100001-110000	9
110001-120000	5
120001-130000	3
150001-160000	1
160001-170000	1
180001-190000	1
340001-350000	1
Grand Total	1248

Table 13.6 shows the population distribution of the tracking kWh savings. The population contains a large number of small projects and a small number of relatively large projects, i.e., the distribution is highly skewed.

Table 13.7: The Skewness of Tracking Savings in the Population

Trk kWh	Projects	Savings
0 - 30,000	85%	45%
30,001 - 100,000	13%	41%
100,001 and up	2%	14%
All Projects	100%	100%

Table 13.7 shows the skewness of savings in the population in another way. The table shows that 85% of all projects have 30,000 kWh or less tracking savings but these projects account for only 45% of all tracking savings. The largest 2% of the projects, those with tracking savings of 100,001 kWh or more, have 14% of all savings. This type of situation is very common in impact evaluation, particularly for commercial and industrial programs. Consequently, stratification by size is generally an extremely important feature of an efficient sample design.

Step 3 – Assess the Error Ratio

Fortunately this program has been offered for several years and was evaluated two years earlier. Although it included short-term end use metering, the prior evaluation used a similar approach to the planned approach. In the prior study the error ratios for gross and net kWh savings were both found to be about 0.4. Since the methodology used in the prior evaluation was comparable and since the program has not changed substantially, an error ratio of 0.4 was assumed to be a good basis for planning the new study.

Choose the Sampling Approach

As noted above, the coefficient of variation of kWh savings was found to be 1.45 and the data collection and analysis costs were expected to be rather high per sample project. Moreover the error ratio was assumed to be 0.4 – much lower than the coefficient of variation. Therefore stratified ratio estimation was clearly preferable to simple random sampling. This decision will be validated in Step 5b.

Step 4 – Assess the Desired Relative Precision

Since stratified ratio estimation is planned, the expected relative precision can be calculated for any given sample size using the equation

$$rp = 1.645 \sqrt{1 - \frac{n}{N}} \frac{er}{\sqrt{n}}.$$

In this application, $er = 0.4$ is the assumed error ratio, $N = 1,248$ is the number of units in the population, and n is the sample size. For example, if $n = 50$, we calculate

$$rp = 1.645 \sqrt{1 - \frac{50}{1,248}} \frac{0.4}{\sqrt{50}} = 0.091.$$

This implies that under the stratified ratio estimation methodology, the expected relative precision would be about $\pm 9\%$ at the 90% level of confidence with a sample of 50 projects.

Table 13.8 shows the expected relative precision for various sample sizes. Figure 13.9 shows these results graphically. From these results, it is clearly practical to plan the new study for $\pm 10\%$ relative precision at the 90% level of confidence.

Table 13.8: Expected Relative Precision vs. Sample Size with Stratified Ratio Estimation

Sample Size	Expected Relative Precision
30	11.9%
35	11.0%
40	10.2%
45	9.6%
50	9.1%
55	8.7%
60	8.3%
65	7.9%
70	7.6%
75	7.4%
80	7.1%

Step 5a – Calculate the Required Sample Size

The preceding results indicate that a sample of about 40 projects is sufficient to provide the desired $\pm 10\%$ relative precision at the 90% level of confidence. To provide a more complete example, we will illustrate the calculations in greater detail. The required sample size can be calculated using the two equations

$$n_0 = \left(\frac{1.645 \text{ er}}{D} \right)^2 \text{ and } n = \frac{n_0}{1 + n_0/N}.$$

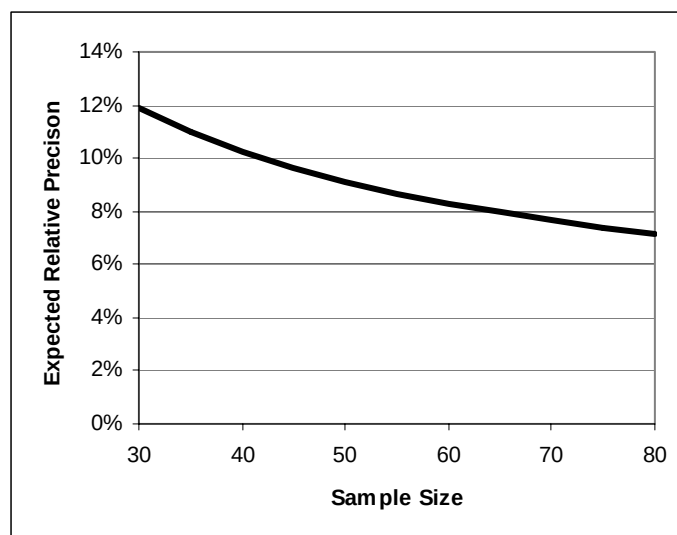


Figure 13.9: Expected Relative Precision vs. Sample Size with Stratified Ratio Estimation

Here $er = 0.4$ is the assumed error ratio, $D = 0.10$ is the desired relative precision at the 90% level of confidence, and $N = 1,248$ is the number of units in the population. Thus we calculate

$$n_0 = \left(\frac{1.645 \times 0.4}{0.1} \right)^2 = 43 \text{ and } n = \frac{43}{1 + 43/1,248} = 42 \text{ projects.}$$

To provide a conservative margin, a sample of 50 projects was planned.

Step 5b – Compare Stratified Ratio Estimation with Simple Random Sampling

This step is not actually required since the assumed error ratio of 0.4, is substantially smaller than the coefficient of variation of 1.45. In order to provide a more complete example, the comparison to the statistical precision and sample size if simple random sampling is used is still provided.

In these calculations the preceding equations are used but with the error ratio replaced by the coefficient of variation, which was found to be 1.45 in Step 2. Table 13.9 shows the expected relative precision for a range of sample sizes. Figure 13.10 shows these results graphically. These results suggest that to provide $\pm 10\%$ precision at the 90% level of confidence, a sample of almost 400 projects would be needed. In other words, the sample would have to be about ten times as large with simple random sampling instead of stratified ratio estimation. These results are expected given that the projects in the population have a highly skewed size distribution and that simple random sampling ignores the tracking estimates of savings.

Table 13.9: Expected Relative Precision vs. Sample Size with Simple Random Sampling

<i>Sample Size</i>	<i>Expected Relative Precision</i>
50	33.0%
100	22.8%
150	18.2%
200	15.4%
250	13.5%
300	12.0%
350	10.8%
400	9.8%
450	9.0%
500	8.2%
550	7.6%
600	7.0%

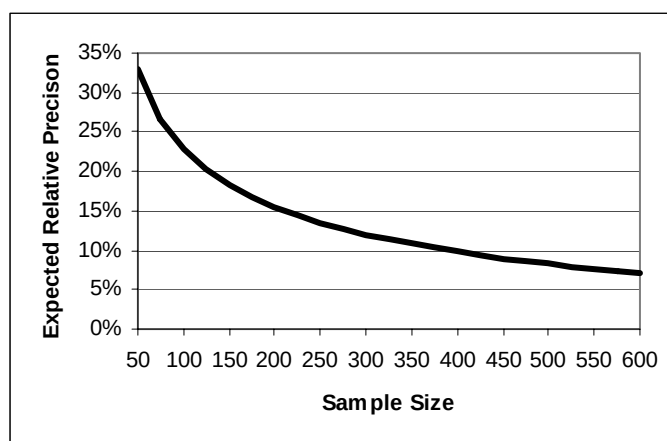
**Figure 13.10: Expected Relative Precision vs. Sample Size assuming Simple Random Sampling**

Table 13.10 shows a simple comparison of the cost of the two approaches. The table shows that under simple random sampling, 390 projects would be needed to provide $\pm 10\%$ precision at the 90% level of confidence. Assuming an M&V cost of \$1,000 per sample project, the total variable M&V cost would be about \$390,000. Under simple random sampling, we have assumed that the statistical sample design and analysis cost would be about \$10,000, for a total variable cost of \$400,000. This neglects the cost of project management, reporting, etc. that can be assumed to be unaffected by the sampling approach.

Table 13.10: Cost Comparison of Simple Random Sampling vs. Stratified Ratio Estimation

Common Parameters	
Desired Relative Precision	10%
Population Size	1,248
Average M&V Cost per Project	\$1,000
Simple Random Sampling Approach	
Coefficient of Variation	1.45
Sample Size before Finite Population Correction	567
Sample Size after Finite Population Correction	390
Total Variable M&V Cost	\$390,000
Statistical Sample Design and Analysis	\$10,000
Total M&V and Statistical Cost	\$400,000
Stratified Ratio Estimation Approach	
Error Ratio	0.4
Sample Size before Finite Population Correction	43
Sample Size after Finite Population Correction	42
Total Variable M&V Cost	\$42,000
Statistical Sample Design and Analysis	\$30,000
Total M&V and Statistical Cost	\$72,000

By contrast, under the stratified ratio estimation approach, the required sample is 42 projects so the total M&V cost would be about \$42,000. Under stratified ratio estimation, the statistical sample design and analysis are assumed to cost \$30,000, for a total variable cost of \$72,000. In this example, simple random sampling is not a feasible choice even though stratified ratio estimation requires more statistical expertise. In other applications, this type of comparison may help determine which approach is preferable.

Step 7 – Construct the Size Strata

Stratification by size is required whenever the projects in the population vary by size. Usually three to five strata are required to provide adequate control over size. In this example five strata will be used since there is substantial variation in the population. Table 13.11 summarizes the assumptions used to construct the strata.

Table 13.11: Assumptions Used to Construct the Size Strata

Sample Size	50
Number of Strata	5
Beta	1
Error Ratio	0.4
Set Gamma	0.8

Table 13.12 shows a portion of the worksheet used to calculate the stratum of each case in the population. The column labeled μ shows the value of $\mu_i = \beta x_i$. The next column shows the value of x_i' . Using these columns, the σ_0 parameter was calculated using the equation

$$\sigma_0 = er \frac{\sum_{i=1}^N \mu_i}{\sum_{i=1}^N x_i'}$$

This provided a calculation: $\sigma_0 = 3.096$. This was used to calculate $\sigma_i = \sigma_0 x_i'$ which is labeled Sigma.

Table 13.12: Constructing the Strata

Project	Trk kWh	μ	x_i'	Sigma	Sort Str	Cum Sigma	Stratum
450	48	3117	22	69	1	69	1
449	75	3056	32	98	2	166	1
453	75	2668	32	98	3	264	1
458	75	14710	32	98	4	362	1
462	75	11142	32	98	5	460	1
108	157	1482	57	177	6	637	1
1107	112,340	51579	10,976	33,976	1238	7,610,012	5
1186	114,621	28496	11,154	34,526	1239	7,644,539	5
1030	114,931	11737	11,178	34,601	1240	7,679,140	5
516	117,148	2000	11,350	35,134	1241	7,714,274	5
987	120,889	129205	11,639	36,029	1242	7,750,302	5
488	121,117	16520	11,656	36,083	1243	7,786,385	5
1242	129,205	44611	12,275	37,998	1244	7,824,383	5
840	152,005	106948	13,979	43,274	1245	7,867,657	5
522	164,464	624	14,889	46,089	1246	7,913,746	5
1032	184,577	18188	16,328	50,545	1247	7,964,291	5
213	345,350	52225	26,953	83,435	1248	8,047,726	5

Then the database was sorted by increasing sigma. The column labeled Sort Str identifies the order of the cases after they were sorted, i.e., the sort order for stratification. Cum

Sigma was calculated using the equation $c_i = \sum_{j=1}^i \sigma_j$. Finally, Stratum was calculated

using the Excel equation $h_i = INT\left(L \frac{c_i}{c_{Ni}} + 0.99999999\right)$.

Table 13.13 shows a summary of the sample design. For example, stratum 1 consists of all projects with tracking savings of 10,128 kWh or less. Stratum 1 contains 694 projects with a total tracking savings of 2,892,887 kWh. The average tracking savings of these projects is 4,168 kWh per project. By contrast, stratum 5 consists of the largest 54 projects. These projects have a total tracking savings of 5,265,839 kWh and an average size of 97,516 kWh per project.

The last column of Table 13.13 shows the desired number of sample projects in each stratum. It is characteristic of this type of design that the sample is equally allocated among the strata. Sometimes, however, the sample design calls for more projects that are in the population in the largest stratum. In this case the sample size is set equal to the number of projects. In other words all projects in this stratum are to be selected with certainty.

Table 13.13: Summary of the Sample Design

<i>Stratum</i>	<i>Tracking Savings</i>				<i>Sample Size</i>
	<i>Projects</i>	<i>Max</i>	<i>Total</i>	<i>Average</i>	
1	694	10,128	2,892,887	4,168	10
2	253	18,981	3,502,474	13,844	10
3	151	35,341	3,979,634	26,355	10
4	96	62,056	4,478,481	46,651	10
5	54	345,350	5,265,839	97,516	10
Total	1,248		20,119,315	16,121	50

Sample Selection

Continuing with the preceding example, Table 13.14 shows a portion of the worksheet used to select the sample. All four steps of sample selection were carried out in this worksheet. Note that the values of “Random” were created using the Excel equation = RAND() and then copied and pasted to replace the equation by the actual values so that they would not be recomputed. Then the cases were sorted according the increasing values of Random and the sort order for selection was recorded (Sort for Selection). The Priority was assigned in each stratum. In each of the five strata, the ten projects with priority 1-10 were designated as the primary sample and the projects with priority 11-20 were designated as backups. If a replacement is required in any given stratum, the first unused project in that stratum should be used as the replacement.

Table 13.15 summarized the characteristics of the primary sample. The table shows the number of accounts in each stratum, the total tracking savings, and the average savings per stratum. The table also shows the total and average savings of all 50 of the sample projects. It should be recognized, however, that the total and average across all strata are not meaningful statistics since the sample projects have not been correctly weighted to reflect the sampling fraction in each stratum.

In each of the five strata, the average savings of the sample projects shown in Table 13.15 should be approximately equal to the average savings of the sample projects shown in Table 13.13. The differences should be due only to sampling variation. In this case, the sample average happens to be smaller than the population average in strata 1 and 4 and larger in the remaining three strata. The overall average size of the sample projects is much larger than the average size of the projects in the population since we have deliberately over-sampled the larger projects. In the analysis, the case weights will be used to avoid any bias in the final results.

Table 13.14: Selecting the Primary Sample and the Backup Projects

Project	Trk kWh	Stratum	Random	Sort for Selection	Priority	Sam Size	Sample Category
55	2,283	1	0.0002	1	1	10	Primary
707	2,632	1	0.0009	2	2	10	Primary
108	157	1	0.0045	3	3	10	Primary
777	978	1	0.0049	4	4	10	Primary
257	157	1	0.0054	5	5	10	Primary
263	6,273	1	0.0057	6	6	10	Primary
418	6,840	1	0.0064	7	7	10	Primary
496	1,577	1	0.0077	8	8	10	Primary
871	5,275	1	0.0102	9	9	10	Primary
113	349	1	0.0113	10	10	10	Primary
462	75	1	0.0130	11	11	10	Backup
1013	7,410	1	0.0162	16	12	10	Backup
1241	2,000	1	0.0186	13	13	10	Backup
503	7,517	1	0.0194	14	14	10	Backup
465	1,786	1	0.0226	15	15	10	Backup
1092	5,045	1	0.0255	12	16	10	Backup
475	4,759	1	0.0259	17	17	10	Backup
879	9,010	1	0.0281	18	18	10	Backup
1152	3,590	1	0.0318	19	19	10	Backup
873	6,180	1	0.0324	20	20	10	Backup

Table 13.15: Summary of the Primary Sample

Stratum	Number	Total	Average
1	10	26,521	2,652
2	10	152,375	15,238
3	10	275,066	27,507
4	10	443,479	44,348
5	10	981,100	98,110
Total	50	1,878,541	37,571

In this example, experienced staff was used to recruit and schedule the sites that were selected for the sample. Table 13.16 shows a portion of the database used to track the recruiting process. The full recruiting database contained all of the available contact information, fields for tracking each attempt to reach the primary contact, and other information used to manage the recruiting process.

The recruiting database listed all 100 selected projects in either the primary or backup sample. The projects were listed in order of priority within each stratum. The status field initially indicated whether the project is in the primary or backup sample. The initial recruiting was limited to the primary sample. When the program participant was contacted, the status field was changed to indicate the result of the contact – in this case either ‘refused’ or ‘scheduled.’ If contact could not be made for any reason, e.g., because of inaccurate contact information that could not be corrected from other sources, the

status was changed to ‘not found.’ Note that in stratum 1, three backups were required to schedule ten projects successfully. The three backups were taken from those on the backup list with the lowest priority, i.e., 11 – 13.

Table 13.16: Portion of the Recruiting Tracking Data

Project	Stratum	Priority	Status	Scheduling Note
55	1	1	Scheduled	
707	1	2	Refused	No lighting installed at actual facility
108	1	3	Scheduled	Meet at Rental Office
777	1	4	Scheduled	
257	1	5	Scheduled	A facility person.
263	1	6	Scheduled	Ring the bell to get let in
418	1	7	Scheduled	
496	1	8	Scheduled	
871	1	9	Not Found	Cannot obtain contact info
113	1	10	Scheduled	Talk to Carrie who will set you up with
462	1	11	Refused	
1013	1	12	Scheduled	Contact will be on vacation so ask for Palmer
1241	1	13	Scheduled	Meet at public buildings dept
503	1	14	Backup	
465	1	15	Backup	
1092	1	16	Backup	
475	1	17	Backup	
879	1	18	Backup	
1152	1	19	Backup	
873	1	20	Backup	

Table 13.17 illustrates some of the data collected in the M&V analysis for the sample projects. For each project in the sample, the following M&V work was carried out:

1. Review the file of each sample project, correct any apparent errors, and recalculate the savings (denoted “Rev kWh”),
2. Visit each sample site, and verify the type and wattage of the lighting equipment actually installed, and recalculate the savings (denoted “Gross Tech kWh”),
3. Verify the quantity of the installed lighting, and recalculate the savings (denoted “Gross Quant kWh”),
4. Assess the hours of use of the lighting in each space, and recalculate the savings (denoted “Gross Hrs kWh”),
5. Interview the program participant, determine whether each measure would have been installed in the absence of the program, and recalculate the savings adjusted for free ridership (denoted “Net kWh FR”),
6. Based on the on-site audit and the participant interview, identify any efficiency measures taken that were not rebated through the program but were the result of the program, and recalculate the savings adjusted for the participant spillover (denoted “Net kWh PSO”).

In this study, similar information was developed for the summer and winter demand savings but this information will not be used in the present example.

Table 13.17: Part of the Sample Database

Project	Stratum	Trk kWh	Rev kWh	Grs Tech kWh	Grs Quant kWh	Grs Hrs kWh	Net kWh FR	Net kWh PSO	Case Weight (w)
55	1	2,283	2,282	2,282	2,282	1,173	1,173	1,173	69.4
108	1	157	157	157	157	158	158	158	69.4
113	1	349	349	349	349	350	350	350	69.4
257	1	157	157	157	157	158	158	158	69.4
263	1	6,273	6,275	6,497	6,960	5,435	5,435	5,435	69.4
418	1	6,840	6,845	6,845	6,845	5,086	5,086	5,086	69.4
496	1	1,577	1,577	1,577	1,577	1,704	1,704	1,704	69.4
777	1	978	978	978	978	981	981	981	69.4
1013	1	7,410	7,409	7,409	7,409	15,486	15,486	15,486	69.4
1241	1	2,000	2,001	2,001	2,001	770	770	770	69.4

Periodic review indicated that recruiting was very successful, with few failures due to bad contact information or refusals by the program participants. Table 13.18 summarizes the recruiting effort. Considering all five strata, there were four failures due to inability to locate a contact person for the site and nine failures due to refusal to participate in the study. In total, 63 sites were contacted to recruit the 50 sample sites, so the response rate was $50/63 = 79\%$. The response rate was at least 71% in each of the five strata.

The recruiting notes were reviewed to understand the reasons for the recruiting failures and whether they might have led to selection bias. There was concern that the sites that were not found might have lower operating hours than the sites that were successfully recruited. If so, the gross realization rate might be overstated.

Table 13.18: Disposition of Sample

Stratum	Not Found	Refused	Scheduled	Total	Response Rate
1	1	2	10	13	77%
2	0	0	10	10	100%
3	1	3	10	14	71%
4	0	2	10	12	83%
5	2	2	10	14	71%
Total	4	9	50	63	79%

Figure 13.11 shows the scatter plot relating the measured gross kWh savings (denoted “Grs Hrs kWh”) versus the tracking estimate of the kWh savings, denoted “Trk kWh”. Each of the 50 sample projects is a separate point in the scatter plot. In this example, the graph shows that one sample project has much lower gross savings than expected from the tracking information. A closer examination revealed that this was project 840. Project 840 was a maintenance garage for a transportation fleet. The tracking system assumed that the lights operated continuously but the on-site audit showed that a substantial portion of the installed lighting was actually operated for about twelve hours a day for five days a week.

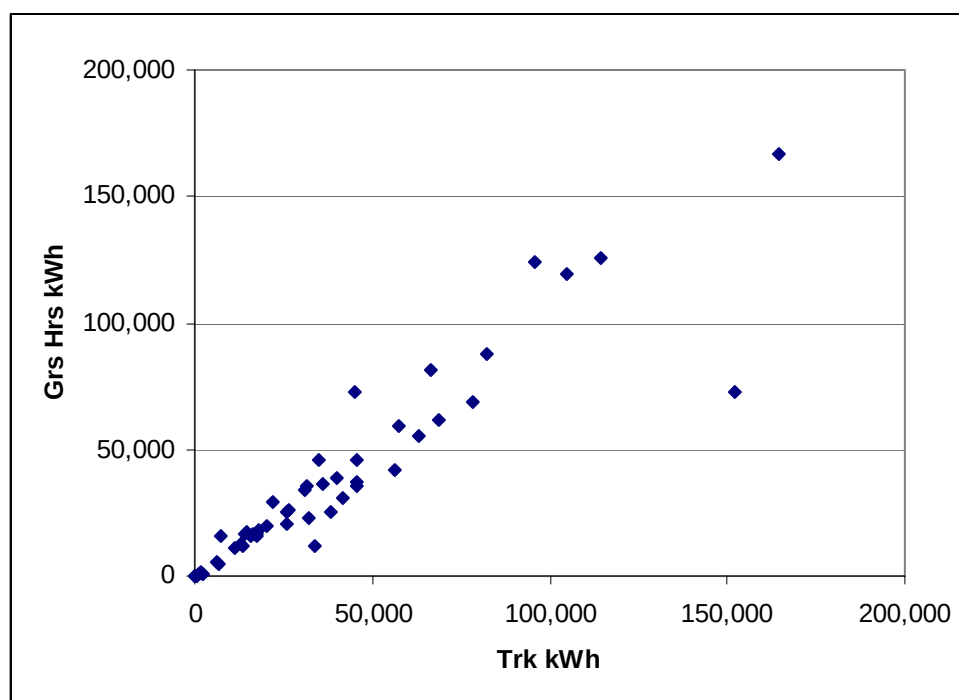
**Figure 13.11: Measured Gross Savings versus Tracking Estimates for Sample Projects**

Table 13.19 shows the ten most influential sample projects as measured by the weighted squared error (WSE) associated with each project. The most influential case is project 1013. This is a small stratum-one project with tracking savings of 7,410 kWh. The

measures savings of this project were found to be 15,486 kWh. After adjusting the tracking savings for the observed realization rate of 0.981, the error was found to be 8,215 kWh. Although this error is relatively small, each project in stratum one is given a high case weight of 69.4. When this case weight is considered, this project makes a much larger contribution of the overall standard error than any other project.

Table 13.19: Using the WSE to Identify Influential Observations

Project	Stratum	Trk kWh (x)	Grs Hrs kWh (y)	Case Weight (w)	w (w-1) e ²	w (w-1) e ²	Sort Influence
1013	1	7,410	15,486	69.4	8,215	320,342,841,430	39
840	5	152,005	72,579	5.4	-76,579	139,338,152,729	36
679	3	33,563	11,704	15.1	-21,230	95,960,582,841	10
204	4	44,967	72,812	9.6	28,687	67,943,968,683	47
550	3	34,944	46,045	15.1	11,755	29,422,331,649	40
677	5	95,833	124,038	5.4	30,000	21,384,148,310	3
1043	4	56,369	42,017	9.6	-13,296	14,595,157,221	6
1036	3	32,037	23,302	15.1	-8,135	14,091,097,698	29
418	1	6,840	5,086	69.4	-1,626	12,546,023,177	17
705	3	22,006	29,209	15.1	7,615	12,347,445,461	25

The second most influential project in the sample was project 840. By contrast with project 1013, this project was the large stratum-five project previously identified in Figure 13.11.

Statistical Analysis

The appropriate statistical analysis will vary depending on the goals and special characteristics of each study. This section will emphasize the methods used to extrapolate the results from the sample to the population and to estimate an error bound for the results. If stratified ratio estimation is used, it is also important to use the sample data to estimate the error ratio to help plan future studies.

Step 1 – Calculate Case Weights

Table 13.20 shows the calculation of the case weights. The case weight is simply the number of projects in the population in each stratum divided by the number of projects in the final sample in the corresponding stratum.

Table 13.21 shows part of the spreadsheet used to assess the gross realization rate. The table shows the ID of each sample project, the tracking savings of the project, the gross savings measured in the M&V analysis, and the case weights determined from the stratum of each sample project.

Table 13.20: Calculating the Case Weights

Stratum	Population	Sample	Case Weight
1	694	10	69.4
2	253	10	25.3
3	151	10	15.1
4	96	10	9.6
5	54	10	5.4
Total	1,248	50	

Table 13.21: Applying the Case Weights to the Sample Projects

Project	Stratum	Trk kWh (x)	Grs Hrs kWh (y)	Case Weight (w)
9	4	45,659	35,606	9.6
48	5	66,718	81,319	5.4
55	1	2,283	1,173	69.4
74	3	31,791	35,392	15.1
108	1	157	158	69.4
113	1	349	350	69.4
188	2	16,408	16,478	25.3
204	4	44,967	72,812	9.6
209	4	45,544	46,204	9.6
257	1	157	158	69.4
263	1	6,273	5,435	69.4
276	2	17,668	16,271	25.3
283	4	38,085	25,451	9.6
418	1	6,840	5,086	69.4
490	4	57,511	59,482	9.6
496	1	1,577	1,704	69.4
522	5	164,464	166,935	5.4
550	3	34,944	46,045	15.1
557	2	17,920	18,243	25.3
608	3	25,843	25,194	15.1
621	5	82,260	87,470	5.4
628	2	14,883	17,602	25.3
660	2	11,220	11,253	25.3
677	5	95,833	124,038	5.4
679	3	33,563	11,704	15.1

Steps 2 & 3 – Calculate the Ratio and the Standard Error

Table 13.22 shows the results for the gross realization rate. Based on the 50 sample sites, the gross realization rate was found to be 98.1%. In other words, the measured gross savings were estimated to be only 1.9% smaller than the tracking savings across all projects in the population. The standard error was found to be 4.6%. This gives an error bound of $\pm 7.5\%$ at the 90% level of confidence. In other words, the 90% confidence interval for the gross realization rate in the population is from 90.6% to 105.7%. The relative precision is 7.7% at the 90% level of confidence.

Table 13.22: Gross Realization Rate

Realization Rate	0.981
Standard Error	0.046
Error Bound	0.075
Low	0.906
High	1.057
Relative Precision	7.7%

The gross realization rate shown in Table 13.22 was calculated from the sample data

illustrated in Table 13.21 using the equation $\hat{B} = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i x_i}$. This equation was converted

into the Excel equation:

$$=\text{SUMPRODUCT}(\$J\$4:\$J\$53,\$I\$4:\$I\$53)/\text{SUMPRODUCT}(\$J\$4:\$J\$53,\$H\$4:\$H\$53)$$

The SUMPRODUCT function in the numerator calculates the sum of the products of the case weights and gross savings. The SUMPRODUCT function in the denominator calculates the sum of the products of the case weights and tracking savings.

Table 13.23 illustrates some of the data used to calculate the standard error of the gross realization rate. For each sample project the error was calculated as $e_i = y_i - \hat{B} x_i$ and labeled as e. Then $w_i(w_i - 1)e_i^2$ was calculated and labeled accordingly. Finally, using another set of Excel functions the standard error was calculated

$$se(\hat{B}) = \frac{\sqrt{\sum_{i=1}^n w_i(w_i - 1)e_i^2}}{\sum_{i=1}^n w_i x_i}.$$

Then the error bound at the 90% level of confidence was calculated using the equation $1.645 se$, and the results were used to calculate the lower and upper boundaries of the 90% confidence interval. Finally, the relative precision was calculated by dividing the error bound by the sample ratio.

Step 4 – Calculate the Error Ratio

The error ratio is the key measure of the association between the y-variable and the x-variable that form the ratio of interest, i.e., between the gross measured savings and the tracking savings of all 1,248 projects in the population. Table 13.24 illustrates part of the spreadsheet used in the calculations. In carrying out this analysis the assumption $\gamma = 0.8$ has been used. This is based on the analysis of many prior evaluation studies.

Table 13.23: Calculating the Standard Error

Project	Stratum	Trk kWh (x)	Grs Hrs kWh (y)	Case Weight (w)	e	w (w-1) e ²
9	2	45,659	35,606	9.6	-9,198	6,984,838,092
48	5	66,718	81,319	5.4	15,850	5,969,398,201
55	4	2,283	1,173	69.4	-1,067	5,407,670,013
74	5	31,791	35,392	15.1	4,197	3,749,693,877
108	5	157	158	69.4	4	62,224
113	1	349	350	69.4	8	299,009
188	1	16,408	16,478	25.3	377	87,442,596
204	2	44,967	72,812	9.6	28,687	67,943,968,683
209	5	45,544	46,204	9.6	1,513	189,063,476
257	3	157	158	69.4	4	62,224
263	1	6,273	5,435	69.4	-720	2,461,033,172
276	2	17,668	16,271	25.3	-1,066	699,122,452
283	3	38,085	25,451	9.6	-11,921	11,732,959,376
418	1	6,840	5,086	69.4	-1,626	12,546,023,177
490	4	57,511	59,482	9.6	3,048	767,084,397
496	2	1,577	1,704	69.4	157	116,430,848
522	4	164,464	166,935	5.4	5,552	732,297,343
550	2	34,944	46,045	15.1	11,755	29,422,331,649
557	4	17,920	18,243	25.3	658	266,534,882
608	2	25,843	25,194	15.1	-165	5,776,173
621	3	82,260	87,470	5.4	6,750	1,082,677,262
628	1	14,883	17,602	25.3	2,998	5,525,726,540
660	4	11,220	11,253	25.3	243	36,334,141
677	4	95,833	124,038	5.4	30,000	21,384,148,310
679	5	33,563	11,704	15.1	-21,230	95,960,582,841

Using this assumed parameter, x_i^γ was calculated. Then the ratio e_i^2/x_i^γ was calculated for each sample project. Then three

$$\text{SUMPRODUCT functions were used to calculate } \hat{e}r = \frac{\sqrt{\left(\sum_{i=1}^n w_i e_i^2/x_i^\gamma\right)\left(\sum_{i=1}^n w_i x_i^\gamma\right)}}{\sum_{i=1}^n w_i y_i}.$$

The result is an error ratio of 0.3. This is smaller than the 0.4 assumed in planning this study.

Table 13.24: Estimating the Error Ratio for the Gross Realization Rate

Project	Stratum	Trk kWh (x)	Grs Hrs kWh (y)	Case Weight (w)	e	x^y	e^2 / x^y
9	2	45,659	35,606	9.6	-9,198	5,341	15,840
48	5	66,718	81,319	5.4	15,850	7,234	34,729
55	4	2,283	1,173	69.4	-1,067	486	2,343
74	5	31,791	35,392	15.1	4,197	3,998	4,405
108	5	157	158	69.4	4	57	0
113	1	349	350	69.4	8	108	1
188	1	16,408	16,478	25.3	377	2,355	60
204	2	44,967	72,812	9.6	28,687	5,276	155,979
209	5	45,544	46,204	9.6	1,513	5,330	430
257	3	157	158	69.4	4	57	0
263	1	6,273	5,435	69.4	-720	1,091	475
276	2	17,668	16,271	25.3	-1,066	2,499	455
283	3	38,085	25,451	9.6	-11,921	4,620	30,763
418	1	6,840	5,086	69.4	-1,626	1,170	2,260
490	4	57,511	59,482	9.6	3,048	6,424	1,446
496	2	1,577	1,704	69.4	157	362	68
522	4	164,464	166,935	5.4	5,552	14,889	2,070
550	2	34,944	46,045	15.1	11,755	4,312	32,047
557	4	17,920	18,243	25.3	658	2,527	172
608	2	25,843	25,194	15.1	-165	3,387	8
621	3	82,260	87,470	5.4	6,750	8,554	5,327
628	1	14,883	17,602	25.3	2,998	2,178	4,126
660	4	11,220	11,253	25.3	243	1,738	34
677	4	95,833	124,038	5.4	30,000	9,665	93,118
679	5	33,563	11,704	15.1	-21,230	4,175	107,946

Step 5 – Estimate the Total Gross Savings of the Program

The results for the gross realization rate were used to estimate the total gross savings of the program as illustrated in Table 13.25. The first entry of the table shows the total tracking savings of all 1,248 projects in the population database. The remaining results were obtained by multiplying the total tracking savings by the results for the realization rate shown in Table 13.22. Then the gross program savings were estimated to be 19,742,490 kWh. At the 90% level of confidence, the total gross savings are estimated to be between 18,225,104 kWh and 21,259,876 kWh.

Table 13.25: Gross kWh Savings of Program

Total Tracking Savings	20,119,315 kWh
Estimated Gross Savings	19,742,490 kWh
Standard Error	922,423 kWh
Error Bound	1,517,386 kWh
Low	18,225,104 kWh
High	21,259,876 kWh
Relative Precision	7.7%

Step 6 – Estimate a Ratio for a Subset

Suppose that the gross savings of the projects implemented by vendor 7 are desired. Table 13.26 illustrates some of the sample data used in the analysis. Note that the indicator variable is the key to the analysis. The indicator is equal to one for vendor 7 and zero otherwise. The tracking and gross kWh are equal to their usual values times the indicator. The rest of the calculation is unchanged.

Table 13.26: Sample Data for Analysis of Vendor 7

Project	Vendor	Stratum	Indicator	Trk kWh (x)	Grs Hrs kWh (y)	Case Weight (w)	e	w (w-1) e ²
9	2	4	0	0	0	10	0	0
48	1	5	0	0	0	5	0	0
55	7	1	1	2,283	1,173	69	-1,430	9,707,856,646
74	7	3	1	31,791	35,392	15	-854	155,446,066
108	7	1	1	157	158	69	-21	2,158,557
113	7	1	1	349	350	69	-48	10,716,611
188	7	2	1	16,408	16,478	25	-2,230	3,056,831,758
204	7	4	1	44,967	72,812	10	21,543	38,315,538,143
209	7	4	1	45,544	46,204	10	-5,723	2,704,004,538
257	7	1	1	157	158	69	-21	2,158,557
263	7	1	1	6,273	5,435	69	-1,717	13,989,736,194
276	2	2	0	0	0	25	0	0
283	4	4	0	0	0	10	0	0
418	4	1	0	0	0	69	0	0
490	4	4	0	0	0	10	0	0
496	7	1	1	1,577	1,704	69	-94	41,897,470
522	7	5	1	164,464	166,935	5	-20,579	10,062,319,626
550	4	3	0	0	0	15	0	0
557	2	2	0	0	0	25	0	0
608	4	3	0	0	0	15	0	0
621	4	5	0	0	0	5	0	0
628	4	2	0	0	0	25	0	0
660	4	2	0	0	0	25	0	0
677	4	5	0	0	0	5	0	0
679	4	3	0	0	0	15	0	0

Table 13.27 shows the resulting gross realization rate and error bound. Table 13.28 shows the estimated total gross savings for vendor 7. Finally, by combining the estimated gross savings shown in Tables Table 13.13.25 and 13.28, it was estimated that vendor 7 was responsible for 29.4% of the gross savings of all projects in the population.

Table 13.27: Realization Rate for Vendor 7

Vendor	7
Realization Rate	1.140
Standard Error	0.097
Error Bound	0.160
Low	0.980
High	1.300
Relative Precision	14.1%

Table 13.28: Total Gross Savings for Vendor 7

Total Tracking Savings	5,811,769	kWh
Estimated Gross Savings	6,626,315	kWh
Standard Error	566,227	kWh
Error Bound	931,443	kWh
Low	5,694,871	kWh
High	7,557,758	kWh
Relative Precision	14.1%	

Analysis of Potential Selection Bias

Recall that in Table 13.18 the response rate was reported as 79%. This present section will summarize a special analysis that was carried out to assess the potential impact of selection bias on the gross realization rate of this program. This analysis was carried out in the following three steps:

1. **Recruiting Case Weights** – create case weights that reflect all 63 projects that were included in the recruiting effort.
2. **Tracking Savings by Recruiting Outcome** – estimate the tracking kWh savings in the program in each of the recruiting categories: not found, refused, and scheduled.
3. **Realization Rate by Recruiting Category** – assess the realization rate in each of the three categories and calculate the overall realization rate of the program.

Table 13.29 shows the development of the recruiting case weights. In these calculations, the sample was considered to be all 63 sites that were contacted during the recruiting work. In each stratum, the case weight was calculated as the ratio between the number of projects in the population and the number of projects in the sample.

Table 13.29: Recruiting Case Weights

Stratum	Population	Sample	Case Weight
1	694	13	53.4
2	253	10	25.3
3	151	14	10.8
4	96	12	8.0
5	54	14	3.9
Total	1,248	63	

Now imagine that the recruiting work could be applied to all of the projects in the population and that each project in the population is put into one of the following three categories: those that would not be found, those that would refuse to participate in the study, and those that would agree to be scheduled. Then the case weights and the tracking kWh savings of each of the 63 projects in the recruiting sample were used to calculate the stratified ratio estimate of the kWh savings of all projects in the population in each of these categories. Table 13.30 shows the results.

Table 13.30: Tracking Savings by Recruiting Category

Stratum	Scheduled	Not Found	Refused	Total
1	1,533,215	288,599	148,102	1,969,916
2	3,950,854	0	0	3,950,854
3	3,139,203	248,143	1,000,057	4,387,403
4	3,700,268	0	775,887	4,476,155
5	3,916,728	557,261	860,998	5,334,986
Total	16,240,268	1,094,003	2,785,044	20,119,315

Finally, the gross realization rate in each category was assessed as shown in Table 13.31. In the scheduled category, the realization rate from Table 13.22 was used, i.e., 0.981. In the Not Found category, it was assumed that the trouble finding the program participant might be because they went out of business or were very seldom in their place of business. This suggests that their operating hours might be shorter than recorded in the tracking system, thereby leading to a lower realization rate than normal. A logical realization rate might be 0.5 or higher for this category. To be conservative 0.5 was assumed here.

In the Refused category, it was assumed that the realization rate was more likely to be similar to that of the Scheduled category. Perhaps program participants refused to cooperate with the on-site audit because they were unhappy with their savings (low realization rate), or perhaps because they were very busy (high operating hours and therefore high realization rate). But it seemed most likely that their refusal was unrelated to their savings. Therefore, the realization rate in the Refused category was taken to be equal to the realization rate in the Scheduled category, i.e., 0.981.

In each of the three categories, the gross kWh savings was estimated by multiplying the estimated tracking savings by the estimated realization rate. Then the total gross kWh savings was calculated across all three categories, 18,606,702, and the overall realization

rate was calculated as 0.925. By calculating the difference between this result and the realization rate previously calculated, shown in Table 13.22, it was concluded that the gross realization rate found from the sample of scheduled projects was likely to have a selection bias of about 0.056 or less.

It is important to recognize that the gross realization rates assumed for the Not Found and Refused categories were extremely subjective since there was very little information for these projects. In particular, there was no objective basis for assuming that the realization rate in the refused category was the same as in the Scheduled category. Sensitivity analysis was used to try different assumed realization rates for the Not Found and Refused categories. At worst, the realization rates could be assumed to be zero in these two categories. In this case the overall program realization rate was found to be only 0.779. In this worst possible scenario, the selection bias would have been 0.202, substantially greater than the error bound reported in Table 13.22.

Table 13.31: Realization Rate by Recruiting Category

	Scheduled	Not Found	Refused	Total
kWh Tracking Savings	16,240,268	1,094,003	2,785,044	20,119,315
Gross Realization Rate	0.981	0.500	0.981	0.925
Est. Gross Savings	15,668,643	638,708	2,299,350	18,606,702

Similar considerations were applied to the net-to-gross analysis, reported below. In this case it was felt that the net-to-gross ratio was unlikely to be substantially different in the three recruiting categories: Scheduled, Not Found, and Refused.

Other Results

The net kWh savings, obtained after adjusting for free ridership, was also estimated. Table 13.32 summarizes the results. Based on the site-specific net-to-gross analysis of the 50 sample projects, the program net realization rate was estimated to be 92.4% $\pm$ 9.4% at the 90% level of confidence. The achieved relative precision for this result was 10.1% at the 90% level of confidence. The corresponding error ratio was 0.41, very close to the 0.4 assumed in planning the study.

Table 13.32: Net Realization Rate

Realization Rate	0.924
Standard Error	0.057
Error Bound	0.094
Low	0.830
High	1.017
Relative Precision	10.1%
Estimated Error Ratio	0.409

The net realization rate for the sample was used to estimate the net savings of the program as illustrated in Table 13.33. From the net savings in Table 13.33 and the gross

savings in Table 13.25, the net-to-gross ratio was estimated to be 18,584,090 / 19,742,490 kWh = 94%.

Table 13.33: Net Program Savings

Total Tracking Savings	20,119,315 kWh
Estimated Net Savings	18,584,090 kWh
Standard Error	1,143,903 kWh
Error Bound	1,881,720 kWh
Low	16,702,370 kWh
High	20,465,810 kWh

Recall that the gross and net savings were measured in a series of steps – file review, technical adjustment for type of fixtures installed, adjustment for the quantity of fixtures installed, the reported hours of use, the reported free ridership, and the reported participant spillover. The analysis of these data is shown in Table 13.34. The table shows each ratio as well as the cumulative product of the ratios up through each factor. In this study, the principle adjustments came from the hours of use (ratio 4) and the free ridership analysis (ratio 5). The participant spillover (ratio 6) partially offset the free ridership. Note that the gross realization rate is the cumulative ratio 4 and the net realization rate is the cumulative ratio 5.

Table 13.34: Decomposition of the Realization Rate into Underlying Factors

	<i>x-variable</i>		<i>y-variable</i>	<i>Ratio</i>	<i>Cum Ratio</i>
1	Trk kWh	versus	Rev kWh	1.000	1.000
2	Rev kWh	versus	Grs Tech kWh	0.998	0.998
3	Grs Tech kWh	versus	Grs Quant kWh	1.003	1.001
4	Grs Quant kWh	versus	Grs Hrs kWh	0.980	0.981
5	Grs Hrs kWh	versus	Net kWh FR	0.941	0.924
6	Net kWh FR	versus	Net kWh PSO	1.018	0.940

These results were also reported in terms of program savings after accounting for each factor, and are shown in Table 13.35.

Table 13.35: Program Savings after Considering Each Factor

<i>Factor</i>	<i>kWh Savings</i>
Trk kWh	20,119,315
Rev kWh	20,120,853
Grs Tech kWh	20,076,578
Grs Quant kWh	20,138,764
Grs Hrs kWh	19,742,490
Net kWh FR	18,584,090
Net kWh PSO	18,909,596

Chapter 14: Evaluation and Cost-Effectiveness

Preface

Most of the chapters in this new Framework provide guidance on evaluation methodologies for conducting program-specific evaluations. This chapter, however, is not providing advice on how to conduct cost-effectiveness analysis. The California Public Utility Commission (CPUC) has directed that the *Standard Practice Manual (SPM): Economic Analysis of Demand-Side Management Programs*, referred to as the *Standard Practice Manual*,³⁵² provide the “how-to” for California cost-effectiveness analysis. In addition, the CPUC has required that a separate, more extensive study be conducted to reexamine the nature and scope of cost-effectiveness testing and the associated issues within the new market structures for California energy resources. For these reasons, this chapter is targeted to three audiences in the area where issues overlap between program evaluations, cost-effectiveness analysis, and their uses and interpretations. First, it will help evaluators to see how their evaluation designs will be used in cost-effectiveness analysis. Second, it will show program staff and administrators who calculate or use cost-effectiveness analysis how evaluation and cost-effectiveness work together. And third, it will help policy makers understand some of the key issues involved in using evaluation results to estimate cost-effectiveness, since these tests are often used to inform a policy decision on whether to continue to invest in a program.

In summary, evaluation results are vital inputs to the cost-benefit tests. Evaluation and cost-effectiveness intersect and interact in several areas, including the following:

- Evaluators must understand the type of data required in order to be able to apply the cost-benefit tests or to structure evaluation efforts that support these requirements;
- The design and delivery of programs can be improved if all parties that are involved in program design, administration, and implementation understand the factors that increase program cost-effectiveness and use these factors to help refine programs; and
- Program implementers should understand temporal and geographical differences in the value of load reductions, so that they can understand why a program that is cost-effective in one time period or area may not be cost-effective at other times or areas.

This chapter highlights issues at the intersection of evaluation and cost-effectiveness analysis and suggests that evaluations might be improved by recognizing these issues. This chapter is written with the following stakeholders in mind: evaluators, policy makers, program designers, program planners, administrators, and those involved in conducting avoided cost studies and cost-effectiveness analysis.

³⁵² (* California State Governor's Office 2001).

Skills Needed

A broad array of professionals is involved in estimating and using estimates of program cost-effectiveness, requiring a broad set of skills across a number of stakeholders. While the issues of cost-effectiveness appear on the surface to be rather simple, comparing the cost of the program-to-program achievements, it is significantly more complicated than this simple comparison of two data points. Professionals who practice in the field of estimating program cost-effectiveness should have at least a sufficient understanding of program theory, economic theory, consumer behavior, energy impact cause and effect relationships, familiarity with energy technologies and related behaviors, an understanding of weather and weather patterns and their influence on energy use and technology efficiencies, and, of course, a strong understanding of the mathematical relationships, theories and tests associated with conducting cost-effectiveness estimations. (Professionals who develop the avoided cost-estimates used to value energy savings in cost-effectiveness analysis need a different set of skills.)

Cost-Benefit Analysis for Evaluators

For analyzing cost-effectiveness, the CPUC uses the standard cost-effectiveness methodologies articulated in the *California Standard Practices Manual (SPM): Economic Analysis of Demand-Side Management Programs*.³⁵³ Two cost-effectiveness tests identified in the SPM are particularly important to the CPUC in evaluating energy efficiency programs on an ongoing basis. The first is the Total Resource Cost (TRC) test – Societal Version. This test, as defined in the SPM, is intended to measure the overall cost-effectiveness of energy efficiency programs from a societal perspective, taking into account benefits and costs from more than just an individual perspective. The CPUC primarily relies upon the results of this test in assessing program cost-effectiveness.

In addition to the TRC test, the CPUC relies on the Participant Test (also identified in the SPM) to evaluate programs that are aimed at inducing individual customers to make energy efficiency decisions. The Participant Test measures the cost-effectiveness of a program from the perspective of energy consumers participating in the program. Proposals for programs designed to provide financial incentives directly to customers are required to include the results of the Participant Test as well as the TRC.³⁵⁴

All resource acquisition programs should have a TRC value greater than 1 to be considered cost-effective. In other words, the present value of the benefits counted under the test must exceed the present value of the costs, so that the benefit-cost ratio is greater than 1.0, or the net benefits (benefits minus costs) are positive.

³⁵³ *Energy Efficiency Policy Manual Version 2*. (* CPUC 2003), page 15.

³⁵⁴ There are other cost-effectiveness tests that are not mentioned in the *Energy Efficiency Policy Manual*. For example, the Utility Costs test focuses on the costs to the utility (and does not include customer costs) and may be used for allocating portfolio resources.

The benefits used in the TRC test are the present values of:³⁵⁵

- The avoided costs of electric and gas energy and capacity, which may be avoided by the participating customer or the utility, depending on future institutional arrangements for electric and gas supply;
- The costs of transmission and distribution investments expected to be avoided;
- The value of the environmental externalities related to burning fuel at the power plant (for electricity) or the end use (for natural gas); and
- The avoided participant capital and operating costs, including reduced water bills and reduced maintenance costs (e.g., less frequent lamp replacements).

The costs used in the TRC are the present values of:

- Direct program implementation and overhead costs, excluding incentives paid to participants;
- The costs of energy, capacity, transmission and distribution, and externalities associated with any increases in energy usage caused by the program;³⁵⁶ and
- The increases in participant capital and operating costs.

Factors Influencing Program Cost-Effectiveness

Important points derived from the above benefit and cost streams that evaluators should understand include the following:

- Benefits are measured by avoided costs, not by energy or peak demand reductions. The true avoided costs of the load reductions will vary with time and location, but these variations may not be fully reflected in the avoided cost stream prescribed for valuing them.
- The TRC test is computed from the present values of benefits and costs. Ongoing costs and benefits are just as important as expenditures in the program year (although future cash flows are discounted).
- Longer-lived measures will have greater benefits than shorter-lived measures, all else being equal.
- From the perspective of the TRC test, a dollar of expenditure by a participant is just as important as a dollar of expenditure by the program implementer.

³⁵⁵ *Standard Practice Manual (SPM): Economic Analysis of Demand-Side Management Programs.* (* California State Governor's Office 2001).

³⁵⁶ Energy use may increase for any of several reasons. Fuel substitution increases the use of one fuel to reduce the use of another fuel. Conservation programs that reduce waste heat in occupied building space will also reduce air conditioning energy use, but will tend to increase space-heating energy use. Load management decreases use at some times, but may increase energy use at other times.

- At the implementation level, the objective is to achieve the maximum improvement in net benefits per dollar cost, rather than to maximize the ratio of benefits to costs. For example, it is better to spend \$50 to save \$100 (net savings is \$50 and B/C is 2.0), than to spend \$10 to save \$30 (net savings is \$20 and B/C is 3.0). Optimization of program design should be driven by the net benefit per application, not the benefit-cost ratio, to avoid cream skimming and creation of lost opportunities.³⁵⁷

Each of the above points has implications for evaluation and program design. As this is an evaluation framework document, some examples of how this could affect evaluation research design are mentioned:

- Because the value of achieved load reduction varies by time and location, at least for larger programs or groups of programs, it is important to calculate and present impact or measurement and verification (M&V) evaluation results by date, time, and geographical location.
- The ongoing costs and benefits included in cost-effectiveness analysis create the need for evaluation to be concerned with the ongoing value of the costs and benefits. Hence, the need for persistence studies for measure retention, technical degradation, and persistence of savings. These have been an important part of the evaluation agenda through the prior measurement and evaluation (M&E) Protocols.³⁵⁸
- In addition to focusing on technologies, it may be important to evaluate other streams of costs or benefits, such as the program's effect on ongoing maintenance costs and periodic training costs as are needed or reduced due to changes made through the program (in equipment or operation). However, in some cases these might be considered non-energy benefits and costs, so see Chapter 11 on Non-Energy Effects Evaluation regarding conducting evaluations to measure these types of benefits and costs. At the same time, if required activities are not undertaken and the consequences mean a shorter measure life or reduced persistence, this assessment needs to be included in the evaluation.
- In calculating the TRC, participant expenditures are just as important as program dollars. Hence, programs need to be designed not only to be cost conscious for the program but also to be cost conscious for participants (including such costs as

³⁵⁷ If budgetary limitations preclude pursuit of all cost-effective efficiency opportunities, the ratio of net benefits to program expenditure is useful in allocating capital among independent efficiency investments (such as towards different types of customers, or on different end-uses or equipment), where deferred savings can be realized later. The B-C ratio is not appropriate among competing investments (on the same end use or equipment), where foregone investment opportunities are forever lost.

³⁵⁸ *Protocols and Procedures for the Verification of Costs, Benefits, and Shareholder Earnings from Demand-Side Management Programs.* (* CADMAC 1999).

training, staffing, maintenance, etc.). This is an area that might best be explored through process evaluation.

- Certain types of costs should be separated from program delivery costs to avoid distorting evaluation findings, as discussed in the next section.

Treatment of Special Categories of Costs

For many programs offered in California over the past few years, some incurred costs are associated with program delivery, but are not necessary for the delivery of program services itself. These costs should be separated from program delivery costs to avoid distorting evaluation findings by including costs that are not costs that would typically be a part of the program. These include, but are not limited to, participant-supported upgrades, program start-up costs, or evaluation costs.

Participant Supported Upgrades

In some program designs, participants may be able to specify equipment for the efficiency improvement that costs more, but offers other benefits that are not associated with improved energy efficiency: a low-flow showerhead with a massage setting, an efficient refrigerator with deluxe finish or customer preferred options, or lighting fixtures that reduce glare and increase productivity. In these situations, it is important to distinguish between the costs associated with the efficiency improvement at the existing level of other amenities (e.g., a white refrigerator replacing a white refrigerator) and the additional costs related to the upgrade (the incremental cost of the chrome finish). In these cases, the cost of the energy efficient upgrade is less than the total cost of the energy efficient technology as installed. The stainless steel ENERGY STAR® refrigerator will cost considerably more than the same technology without the stainless steel surface. If the cost of the stainless steel unit is considered as part of the cost of the energy efficient upgrade, then the cost-effectiveness assessment includes costs that are not related to the energy efficient improvement. When possible, these costs should be excluded, unless the program theory indicates that the stainless steel exterior is a program component that adds to the energy efficient improvement and is, therefore, a part of the program offering.

Participants may coordinate substantial remodeling with efficiency upgrades to lighting, HVAC and other systems. The costs of other activities, even if triggered by the efficiency measures and implemented through the same contractor, are not part of the measure cost. Where possible, the change in energy use should only be for those measures that are being replaced or retrofitted that directly relate to the program theory of how the technology is providing savings. On the other hand, if one were to include the energy savings from these other measures, then the costs for these additional measures should be calculated as well.

Start-Up and Future Improvement Costs

In evaluating and interpreting the cost-effectiveness of programs, it is important to distinguish between costs incurred in the past (particularly non-recurring “start-up costs”) and costs that would be incurred for a similar program in the future. For example, non-repeating start-up costs such as market research, protocol development, regulatory costs associated with program approval, and contracting are different than normal program operating costs. Similarly, the incremental costs of gathering data for future improvements and future programs (e.g., market analysis), beyond those required for the operation and reporting of the existing program, are also costs that should not be considered normal program operational costs.

Charging these costs to the current year of program operation might make the program appear to be uneconomic, even if the program was producing benefits well in excess of normal direct program costs. All costs must be reported and included in historical summaries, but costs unrelated to continued operation should be excluded from the analyses of the cost under *steady-state conditions* or to determine the cost of *continuing* the program.

This concept is already included in key California cost-effectiveness policy documents. The CPUC recognizes that start-up costs are non-repeating, and should not bias analysis of continuing programs:

“The TRC should be calculated by treating programs as multi-year (rather than single-year) activities so that programs explicitly designed as integrated, multi-year strategies, which may have modest benefits (and/or high start-up costs) in early program years, could be evaluated considering the expected larger benefits (and/or lower costs) in later program years.”³⁵⁹

Evaluation Costs

As a general philosophy, evaluation costs can be considered as a standard cost of running an efficiency program with Public Goods Charge funding. At the same time, the level of evaluation costs could swing significantly from year to year depending upon the life cycle of the program, market circumstances, and purpose of the evaluation. This can create situations where use and interpretation of including these costs in a cost-effectiveness test could be problematic.

An extreme example could be found in a pilot program testing a new program concept where the evaluation costs could be as high, or higher, than the rest of the program costs. A continued program would be expected to have far lower evaluation costs than what might occur on a periodic schedule. Including the pilot’s evaluation costs in the cost-benefit analysis that is used to decide whether to continue this program to full operation would make no sense and doom the pilot to failure.

³⁵⁹ *Energy Efficiency Policy Manual Version 2. (* CPUC 2003).*

In Chapter 4 on Evaluation Overview and Issues the concept of formative evaluations versus summative evaluations is discussed. To briefly recap: in summative evaluations, the focus is on documenting and measuring the effects of a program; in formative evaluations, the focus is on understanding why those effects occurred and identifying ways to improve a program. As a policy and evaluation tool, cost-effectiveness analysis could also conceptually be divided into formative and summative. If a test of cost-effectiveness is being used to inform a policy decision of whether to continue to invest in a program or program concept, it is best if the cost-effectiveness analysis take a formative perspective.

Given these concerns, good practice seems to indicate that program cost-effectiveness should be analyzed and reported both ways. Cost-effectiveness analysis needs to be conducted with all relevant program costs (to include apportioning administrative costs from the implementing organization as is relevant to the administrative needs of the program, regulatory costs, and evaluation costs), as provided by the *Standard Practice Manual*. Cost-effectiveness analysis could be reported without regulatory and evaluation costs. In the reporting of these cost-effectiveness tests, including a qualitative discussion about recurring future costs versus one-time or unusual costs would also add to the quality and usefulness of the reporting. This provides administrators and policy makers the best information for both summative and formative purposes.

Excluding evaluation or regulatory costs from program-specific cost-effectiveness tests does not in any way mean that these costs should be excluded in an assessment of the cost-effectiveness of the entire Public Goods Charge process for designing and delivering energy resources to the State of California. Rather, this concept is included here to indicate that evaluation costs that are above and beyond the costs that are needed to acquire the resources might be included in a cost-effectiveness assessment of the program portfolio, but not necessarily included in the assessment of individual programs.

Understanding Temporal and Geographical Variations in Avoided Costs

The value of a kilowatt-hour of electricity, or a therm of natural gas, varies over time and space.

- Avoided electric energy costs vary with load level. When demand is high, market prices for energy also tend to be high. As a result, market energy prices are typically higher during weekday business hours (“peak hours”) than on weekends, and lower late at night (“off-peak hours”). Energy prices are higher in months with high cooling and heating loads and lower in milder months when the demand for energy is lower.

- Higher electric loads result in higher energy losses in the transmission and distribution systems, compounding the variation in energy prices at the generation level.
- Natural gas prices tend to vary from day to day (rather than hour to hour, as for electricity), with higher prices on the cooler days that require more heating, and hence more gas.
- Energy costs vary from one part of the state to another, depending on transmission constraints.³⁶⁰
- The costs of transmission and distribution (T&D) capacity to deliver energy to consumers are driven by peak loads. Those peaks may occur at different times (and even different seasons) for various T&D components.³⁶¹

One important implication of the variation of avoided costs over time is that various load shapes of energy savings will have different benefits per kWh (or per therm). Reduction in electric loads for industrial motors that run all the time will save relatively little T&D capacity per kWh, and a large fraction of the energy will be saved at relatively low-priced hours. Reductions in commercial lighting loads will tend to be concentrated during business hours, with higher average energy prices, and more capacity savings per kWh. Reductions in commercial cooling loads in a summer peaking area will usually save energy primarily in the highest priced hours, when the T&D system is most heavily loaded, which creates high avoided costs per kWh and per kW.³⁶²

These variations cannot be entirely captured by dividing avoided costs into time periods by the calendar or clock. Within any particular period, such as summer weekday afternoons, loads, energy costs and losses will be higher on some days than others, depending primarily on weather conditions. Load reductions that tend to follow load (such as air conditioner efficiency) will have higher avoided costs than reductions that follow the clock. In addition, capacity availability can also influence cost. Generation facilities may be closed or in reduced capacity mode because of a range of conditions. Generation facilities may be shut down for scheduled or unscheduled maintenance causing shortages that increase cost. It is possible that the value of energy saved via a resource acquisition program can be significantly more valuable if it is saved during times of limited supply. These conditions mean that the value of energy saved from an energy program depends on the price of the energy at the time it is being saved.

³⁶⁰ Whether those differences are large and persistent enough to be reflected in avoided costs for program valuation remains to be seen.

³⁶¹ While avoided electric T&D costs vary widely from one substation or feeder to the next, developing and using avoided T&D costs on that level of detail (potentially thousands of different values and peak periods across the state) is not likely to be feasible for most purposes. Whether aggregate regional estimates of avoided T&D costs would be more accurate than a statewide average, or vary enough to be worth reflecting in avoided costs for program valuation, remains to be seen.

³⁶² Even for a single end use, different measures may have different load shapes and hence different avoided costs. For example, daylighting sensors on commercial lighting are likely to preferentially reduce lighting load at high-priced on-peak summer midday hours, while occupancy sensors may primarily reduce loads at night.

Reductions affecting the peak of the average system load shape will tend to have higher avoided costs than reductions that are reducing energy use for all hours, since energy prices, line losses, and stress on the T&D system tend to be high when loads are high. Evaluators must be able to not just report how much energy is being saved, but be able to identify the time period over which the energy is saved and the amount of energy that is being saved within relevant price periods in order to identify the avoided costs associated with a program or group of programs.

Recent changes in energy demand, changes in the regulatory environment, and changes within the energy supply and distribution environment have had, and are expected to continue to have, significant impacts on the price of energy in California. As a result, there is a significant amount of uncertainty associated with projecting the avoided cost associated with resources generated from energy efficiency programs. In the future, it may be more appropriate to use avoided cost projections that are reported with bounded confidence intervals and relative precision levels, so that the level of uncertainty associated with any given point estimate can be used to identify the expected and possible cost-effectiveness range associated with energy programs. (The methods described in Chapter 12, Uncertainty, can be used for chaining of results to estimate an overall relative precision.)

At the present time, the avoided cost calculation methods used in California do not incorporate an *economic condition hedge factor*. This would allow the avoided cost equation to include an estimate of the value of energy efficiency resources as a hedge against the economic effects of blackouts and curtailments on the California economy. When Public Goods Charge funds are used to support energy programs, there may be a need to incorporate a factor in the avoided cost estimate that incorporates the value associated with the effects of limited supplies on the California economy and the benefits resource acquisition programs can provide toward reducing the economic impacts of these conditions.

Providing Useful Avoided-Cost Estimation to Evaluators

The derivation of avoided costs is beyond the scope of this report. This section considers the form in which the estimated avoided costs should be provided to evaluators, to promote efficient and consistent use of those estimates. A number of features would contribute to making avoided costs easier for evaluators to apply, and to making the resulting program cost-effectiveness analyses more accurate. These comments do not presuppose the methodology used to produce the avoided costs.

First, the avoided costs provided to evaluators should be as simple and as complete as possible, so that evaluators need not be concerned with more separate cost components

than necessary. For example, the avoided costs should include to the extent possible all externalities and (for electricity) load-specific losses to the end use.³⁶³

Second, avoided costs should be available for common technology-associated load shapes (e.g., commercial chilling, commercial lighting, and residential air conditioning) as well as for constant loads within each rating period. These avoided costs typically vary between geographical regions due to climate and due to differing capacity, demand, transmission and distribution constraints.

Third, avoided costs should be available in various levels of detail, reflecting the range of program complexity and sophistication of potential administrators. At one extreme, for complex custom measures for large customers (such as large commercial new construction programs, or comprehensive industrial process retrofits), which are likely to be modeled and/or metered in great detail, avoided costs should be available as hourly demand and energy values, so that load effects by time and weather conditions can be matched with avoided costs. For less complicated applications, the avoided energy costs should be listed by rating period, weighted by the hourly loads of various end uses and load types, while avoided demand costs should be listed, with the times and weather conditions driving the peak demands clearly identified.³⁶⁴ These avoided costs can be provided on an annual basis, with the discount rate specified,³⁶⁵ or as a present value for various lifetimes.

Fourth, avoided peak demand energy savings are likely to be based on avoided purchased energy during this time period. These savings should actually reflect the system benefits of reducing peak demand (e.g., less purchase of capacity, less spinning reserve, etc.).

Finally, for small programs using well-characterized measures, it would be helpful to have all the avoided costs for each common measure (e.g., residential lighting, small-commercial refrigeration) reduced to a single value of present value per annual MWh (or MMBtu), which includes load-weighted energy, demand, externalities and loss costs.

³⁶³ If the avoided costs include avoided losses only to the customers' meter, the program evaluation will need to estimate losses on the customer's side of the meter, especially for customers metered at primary voltage.

³⁶⁴ For example, the peak conditions for a geographical area might be listed as "September weekday 2 PM, 95° F."

³⁶⁵ In evaluating energy efficiency program proposals, the CPUC requires everyone to use a pre-established discount rate of 8.15% (as noted in the *Energy Efficiency Policy Manual*). (* CPUC 2003).

Chapter 15: Overarching Evaluation Studies

In addition to the need for evaluating specific programs, or groups of related programs, there are significant benefits from conducting “overarching studies” to collect and analyze information at a statewide or regional level that can be used for program and portfolio evaluation planning as well as policy decision-making. While some of the overarching information can be collected in evaluations of individual programs, large-scale studies may be needed that encompass a broader array of programs. These evaluation approaches are not addressed in this Framework. However, these overarching studies are important for understanding the market in which energy programs are placed and the types of measures that could be considered for inclusion in programs and the level of savings that can be achieved from programs and program-induced actions. The types of overarching studies include, but are not limited to:

1. Measure saturation studies,
2. Energy-savings potential studies (technical, economic, achievable-market),
3. Portfolio analyses (including “best practices” and “lessons learned” studies),
4. Market and market operations analysis (beyond program level),
5. Studies that update key parameters that influence multiple programs (e.g., measure lifetime, avoided costs), and
6. Development of improved methodologies for evaluating programs.

The CPUC, CADMAC and CALMAC³⁶⁶ have funded these and other studies in the past, and the CPUC has been requiring these studies in 2003 and 2004 to help understand and plan for California’s energy supply options today and for the future. The following paragraphs provide a summary review of these types of studies and indicate their importance in planning the needs of California’s energy future and the amount of energy available via energy efficiency programs.

Measure Saturation Studies

It is important to establish a credible platform from which to determine which types of energy efficiency programs should be offered, which sectors should be targeted and what technologies should be included in California’s programs. Measure saturation studies help to achieve this objective by collecting and analyzing data on the types of measures that are, for example, installed in residential and commercial buildings and identifying their current level of energy efficiency. Ideally, these types of studies should be conducted annually on all key energy efficiency measures. However, due to budgetary concerns and in some cases the pace of normal market change, it is more realistic to expect these studies to be conducted periodically over time. Or this effort can focus on

³⁶⁶ CPUC – California Public Utilities Commission; CADMAC – California Demand-Side Management Measurement Advisory Council; CALMAC – California Measurement Advisory Council.

conducting annual or bi-annual studies in which the studies focus on just those technologies in which significant change is expected because of recent market changes or program accomplishments. Alternatively, saturation studies can be rotated through a series of key energy-intensive sectors, or energy-intensive end uses. Examples of past saturation studies in California include the following by: ADM Associates and TecMRKT Works,³⁶⁷ Quantum Consulting,³⁶⁸ RER,³⁶⁹ RLW Analytics,³⁷⁰ and SDG&E.³⁷¹

The primary purpose of the measure saturation study is to baseline the status of the market in terms of the current penetration of energy consuming technologies or behaviors. This is done so analysts and policy makers can document changes in the market and identify those parts of the market in which energy savings potentials can be achieved through future programs and technology offerings.

Energy-Savings Potential Studies

In addition to having a good saturation baseline it is important to know where one should be headed with regards to markets and technology selection so that the appropriate energy efficiency programs can be identified and later designed and implemented. This is particularly important for resource acquisition programs. One tool for identifying the goals for a portfolio of programs is energy-savings potential studies. There are three main types of potential studies: technical potential studies, economic potential studies, and achievable potential studies.³⁷²

Technical potential is defined as the complete penetration of all measures analyzed in applications where they were deemed technically feasible from an engineering perspective. That is, the population of customers has facilities and technologies in which the energy efficient technologies or behaviors *can* be used.

Economic potential refers to that part of the technical potential that can *cost-effectively* be obtained when compared to supply-side alternatives. As technology and energy prices change due to changing market conditions, including deregulation or other events, the levels of economic potential also change.

The *achievable potential* refers to the amount of savings that can occur in response to specific program designs and delivery approaches, including program funding and measure incentive levels. Achievable potential studies are sometimes referred to as market potential studies as they estimate what can be obtained within a specific set of

³⁶⁷ *Statewide Survey of Multi-Family Common Area Building Owners Market: Volume 1: Apartment Complexes.* (ADM Associates and TecMRKT Works 2000).

³⁶⁸ *Statewide LED Traffic Signal Saturation Study.* (Quantum Consulting Inc. 2001).

³⁶⁹ *Direct Assistance Market Saturation Study.* (Regional Economic Research 1991).

³⁷⁰ *Statewide Residential Lighting and Appliances Saturation Study.* (RLW Analytics 2000).

³⁷¹ *Commercial and Industrial Lighting Retrofit Program: Base Equipment Saturation and Operating Hours by Building Type.* (San Diego Gas & Electric Company 1991).

³⁷² *California's Secret Energy Surplus: The Potentials for Energy Efficiency.* (Rufo and Coito 2002).

program design and implementation conditions presented to the target market. The achievable potential is what can be achieved through market interventions. It should be noted that the achievable or market potential is not fixed and is not easily estimated. The achievable or market potential is highly dependant on the programs offered, the technology or technology-mix presented, and especially the characteristics of the program relative to the approach used to reach the target markets, offer program services, and other program design and presentation factors. Over the years, policy makers have tended to rely on economic potential projections and the amount of energy that can be acquired before program costs make the acquisition no longer cost-effective. This type of assessment examines different program funding levels and program design approaches using market penetration assumptions to arrive at an estimate of what can be cost-effectively achieved. However, these types of estimates are often plagued by untested assumptions of what is required to obtain a given level of penetration within a specific market sector or market segment. As more process and market evaluations are conducted, policy makers are able to develop better, more accurate assumptions about what is needed from a program to achieve a given level of penetration.

The following diagram is presented in the *California's Secret Energy Surplus* report³⁷³ and demonstrates the relationship between the market's technical, economic and achievable potential as well as the level of naturally occurring energy efficient change that occurs in the market.

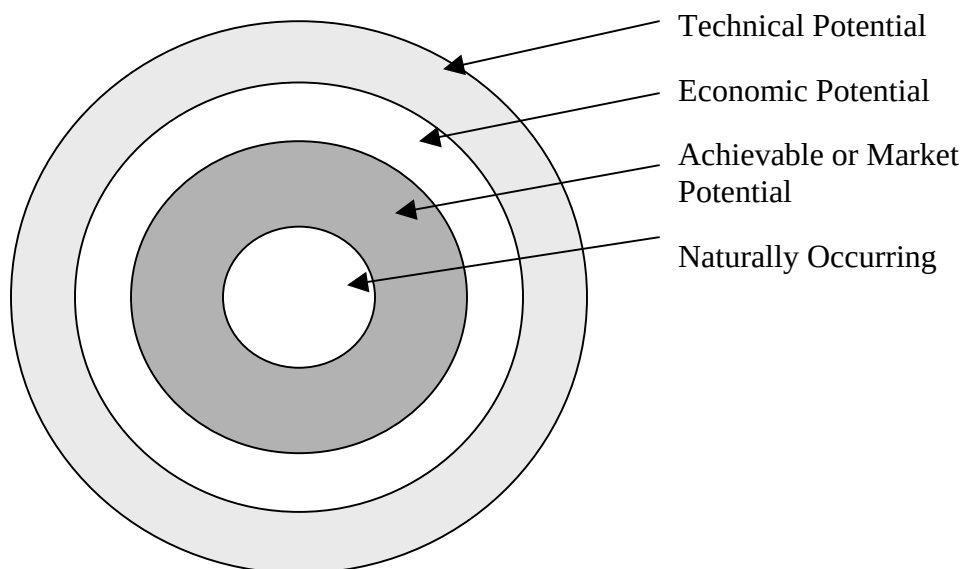


Figure 15.1: Conceptual Relationship Among Energy Efficiency Potential Definitions

These studies can be important decision tools as reflected in the 2003 CPUC rulemaking that requires the CPUC to establish goals towards achieving the energy savings

³⁷³ *California's Secret Energy Surplus: The Potentials for Energy Efficiency.* (Rufo and Coito 2002).

available.³⁷⁴ Examples of past savings potential studies in California include work by Kema-Xenergy,^{375, 376, 377, 378a, 379} and Rufo and Coito.³⁸⁰

Persistence and Retention Studies

Persistence and retention studies are a class of evaluation studies that assess the long-term persistence of energy savings and effective useful life (EUL) of measures installed due to energy efficiency interventions. Persistence of energy savings is a combination of measure retention, measure installations retained versus those that have failed or that have been removed, and the incremental technical degradation (the difference between the technical degradation of the efficient equipment versus the degradation of standard equipment). The EUL, as defined by the prior measurement and evaluation (M&E) Protocols,³⁸¹ is “an estimate of the median number of years that the measures installed under the program are still in place and operable.”

The prior M&E Protocols had a required schedule for periodic program-specific retention and EUL assessment evaluations to be performed. Given this, there have been a large number of these studies conducted in California. (See the CALMAC searchable database at <www.calmac.org>.) The purpose of these studies was to estimate measure retention and assess whether ex-ante EUL appeared reasonable to retain (within the statistical confidence of the estimates or if it appeared to be the most reasonable assumption given unreasonable EUL ex-post estimates) or whether analysis provided the information to estimate an ex-post EUL that differed from the ex-ante EUL. Examples include studies done by Quantum Consulting (residential refrigerators),³⁸² Megdal & Associates (residential weatherization),^{383, 384} ASW Engineering (commercial gas equipment, primarily cooking),³⁸⁵ RLW (non-residential new construction),³⁸⁶ Dohrmann

³⁷⁴ Assigned Commissioner’s Ruling Proposing Direction and Scope for Further Rulemaking. (CPUC 2003).

³⁷⁵ *Conservation Potential Study, Vol 1., Results and Methods.* (Kema-Xenergy 1992a).

³⁷⁶ *SCE Energy Efficiency Potential Study.* (Kema-Xenergy 1992b).

³⁷⁷ *Commercial Sector Energy Efficiency Potential Study.* (Kema-Xenergy 2002).

³⁷⁸ *California Statewide Residential Energy Efficiency Potential Study.* (Kema-Xenergy 2003b).

³⁷⁹ *California Statewide Commercial Sector Natural Gas Energy Efficiency Potential Study.* (Kema-Xenergy 2003a).

³⁸⁰ *California’s Secret Energy Surplus: The Potentials for Energy Efficiency.* (Rufo and Coito 2002).

³⁸¹ *Protocols, and Procedures for the Verification of Costs, Benefits, and Shareholder Earnings from Demand-Side Management Programs* (* CADMAC 1999).

³⁸² *Fourth Year Retention Study for PG&E’s 1996 & 1997 Residential AEI Program Refrigeration Technology.* (Quantum Consulting Inc. and Megdal and Associates 2001).

³⁸³ *Final Report: Statewide Study of the Retention of Measures Installed Under the Direct Assistance Program.* (Megdal & Associates and ASW Engineering Management Consultants Inc. 1998).

³⁸⁴ *Final Report: Measure Retention Study -- 1994 & 1995 Residential Weatherization Programs (RWRI).* (Megdal & Associates and ASW Engineering Management Consultants Inc. 1999)

³⁸⁵ *Measure Retention Study of the 1996 Commercial Energy Efficiency Incentive (EEI) Program.* (ASW Engineering Management Consultants Inc. and Megdal & Associates 2001).

³⁸⁶ *SCE Non-Residential New Construction Persistence.* (RLW Analytics 1998).

(commercial, industrial, and agricultural equipment),³⁸⁷ Richardson (residential new construction),³⁸⁸ and Bordner (various).³⁸⁹ An alternative method of estimating measure lifetime could be a study outside of a program that used a field study of age at replacement, as seen in a study done by Robison et al.³⁹⁰

Several overarching technical degradation studies were sponsored by CADMAC in the late 1990s as being a more cost-effective way to conduct these than through program-specific or even technology-specific studies. An article discussing the issues surrounding technical degradation and early survival analysis can be found in O'Drain and Caulfield 1998.³⁹¹ Most of these studies were conducted in the late 1990s (e.g., Proctor Engineering Group,^{392, 393, 394, 395, 396} and SBW Consulting³⁹⁷).

Survival analysis is the process of analyzing empirical failure/removal data in order to model a measure's survival function. A variety of types of survival analysis was used in the California retention studies to date. Besides finding relatively high retention rates in most cases, a consistent finding in these prior studies is that a longer period of time is needed for conducting these studies so that technology failure and removal rates can be better documented and used to make more accurate savings persistence and persistence impact models. This suggests that future evaluation policies should significantly alter the timing of retention and persistence studies.

As mentioned earlier, the prior M&E Protocols required program-specific retentions studies at specific times. Given the experience gained through these studies, a more cost-effective alternative might be to conduct retention studies by sector, across numerous programs and program years as overarching or consolidated evaluation studies. Though the retention analyzed could still be program-specific, conducting the data collection and analysis across a sector or across multiple sectors should provide a more cost-efficient alternative. Additionally, for larger commercial and industrial customers that can have

³⁸⁷ "A Longitudinal Study of Non-Residential DSM Measure Retention." (* Dohrmann et al. 1999).

³⁸⁸ "Measure Retention in Residential New Construction." (* Richardson and Skumatz 2000).

³⁸⁹ "The Application of Survival Analysis to Demand-Side Management Evaluation." (* Bordner et al. 1994).

³⁹⁰ "Measure Lifetime Derived from a Field Study of Age at Replacement." (* Robinson et al. 1996).

³⁹¹ "Assessing Persistence: Experiences Documenting Savings Persistence Under the California Protocols." (* O'Drain and Caulfield 1994).

³⁹² *Summary Report of Persistence Studies: Assessments of Technical Degradation Factors, Final Report* (Proctor Engineering Group 1999).

³⁹³ *Persistence #3A: An Assessment of Technical Degradation Factors: Commercial Air Conditioners and Energy Management Systems, Final Report.* (Proctor Engineering Group 1999a).

³⁹⁴ *Statewide Measure Performance Study #2: An Assessment of Technical Degradation Factors, Final Report.* (Proctor Engineering Group 1998b).

³⁹⁵ *Negative Technical Degradation Factors Supplement to Persistence Studies, Final Report.* (Proctor Engineering Group 1998a).

³⁹⁶ *Statewide Measure Performance Study: An Assessment of Technical Degradation Rates, Final Report.* (Proctor Engineering Group 1996).

³⁹⁷ *Pacific Gas & Electric Company PY94 Nonresidential New Construction Retention Study.* (SBW Consulting Inc. and Ridge & Associates 1999).

participants in multiple programs across multiple years, this consolidated sector or cross-sector approach may be less intrusive to the participants.

Portfolio Analyses

This Framework is defining portfolio analyses as a class of evaluation studies that work with, evaluate and assess the portfolio of Public Goods Charge programs, as well as other energy efficiency, demand-response, procurement, and renewable energy efforts.

The most common type of portfolio analysis, and the one the energy efficiency community is most familiar with in California, is one that aggregates energy and demand savings (a summative evaluation) with an effort to ensure no double counting of effects. This is done to estimate the PGC funded total energy and demand savings. This estimate can be used to compare to the CPUC's energy efficiency goals as a way of measuring progress and can be used to inform power-planning efforts. The analysis could also be conducted for key market segments (e.g., residential, commercial, and industrial), or key end uses (lighting, motors, HVAC, etc.) or for the programs offered by a single administrator or group of administrators (e.g., PG&E's portfolio or third-party administrator X's portfolio) or for a given area (e.g. Orange County portfolio), or a portfolio of programs by size (e.g. programs that spend less than X dollars per year or acquire less than X kWh per year) or as is primarily used in this Framework, the Public Goods Charge portfolio, or perhaps in the future, the resource procurement portfolio. This type of evaluation could assess overall portfolio level cost-effectiveness, including all costs for all program types (resource, information, demonstrations, resource centers, infrastructure development). An example of this type of portfolio evaluation is the *Summary of 2001 Programs* completed by Global Energy Partners.³⁹⁸

A summative portfolio evaluation can be relatively simple or quite complex. For example, the analysis of the portfolio of programs could be a simple aggregation of the program and evaluation data collected on each program (costs, savings, number of measures, number of participants, etc). Nonetheless, techniques need to be incorporated to assess instances of double counting and minimize these occurrences. The methods needed to ensure there is no (or a limited amount of) double counting gets significantly more complex if there are both market effects estimates and resource acquisition estimates as part of the program effects being addressed. A similar but opposite concern that would add complexity to this type of evaluation would be when additional market transformation evaluations are conducted to provide input into a summative Portfolio evaluation. These efforts would be undertaken in order to measure market effects created by the aggregation of programs that were not a component of any other program's or set of programs' market transformation program evaluation (i.e., adding in savings missed from study aggregations that "fell through" the accounting cracks).

³⁹⁸ *California Summary Study of 2001 Energy Efficiency Programs*. (Global Energy Partners 2003).

The above analysis is primarily a “bottom-up” assessment, adding up the program results and then working at minimizing overlap and filling gaps in that analysis. An alternative summative portfolio evaluation could be performed that would be a “top-down” approach. This could be done through evaluations of statewide energy usage, trends, changes, and the causes of these changes, one of which is the program interventions. Another portfolio evaluation could include matching, comparing, and assessing both approaches together in order to triangulate to an estimate that might help ensure the estimate is not biased. (Of course, this requires having studies that produced both types of estimates.)

An economic impact evaluation of program-induced efficiency investments is a type of non-energy effects evaluation. (See the Chapter 11 on Non-Energy Effects Evaluation for this Framework’s decision roadmap concerning assessing these effects in California.) These types of studies are often conducted at a sector or portfolio level and therefore may be viewed as a type of portfolio evaluation. Economic impact evaluations of portfolios of programs are being done in New York³⁹⁹ and Wisconsin,⁴⁰⁰ and have been done for Austin Electric.⁴⁰¹

Another type of related portfolio analysis that could be both summative and formative is analyzing the portfolio from a systems analysis approach and examining synergies between programs. Synergies could be either positive or negative. The analysis could be done to document synergies (summative) or as an evaluation to assess how positive synergies can be increased and better leveraged. This type of work has begun in New York.⁴⁰²

There are also several types of formative portfolio evaluations that could be undertaken. One of which would be to evaluate the portfolio within the financial perspective of the portfolio. In other words, to evaluate the portfolio in such ways as to obtain a portfolio risk assessment. This would include examining the tradeoffs between resource acquisition and market transformation, high risk and low risk programs/target markets, and pilot, equity, or technology development support versus established acquisition programs. Finding the proper mix of investments to maximize savings at an appropriate risk level for PGC-funded or procurement-funded investments would not be a small task. Though the concept has been discussed,⁴⁰³ we know of no actual formative portfolio evaluation of this type that has been conducted in the energy efficiency field.

³⁹⁹ “Linking Market-Based Energy Efficiency Programs to Economic Growth, Sustainable Development and Climate Change Objectives.” (Smith et al. 1998).

⁴⁰⁰ “Quantifying Economic and Environmental Benefits.” (Sumi et al. 2002).

⁴⁰¹ “The Development of a Local Energy Efficiency Economic Impact Model for Use in Integrated Resource Planning.” (Megdal and Rammaha 1992).

⁴⁰² “A Theory-Based Systems Approach for Evaluating Energy Efficiency Investments Portfolios.” (DeCotis and Munro 2001); “Systems-Based Portfolio Evaluation: Diagnostic Benefits and Methodological Challenges.” (DeCotis et al. 2002).

⁴⁰³ “Portfolio Approach to Designing and Evaluating Buildings Energy Efficiency Programs.” (DeCotis et al. 2000).

A more straightforward formative portfolio evaluation would involve combining potential studies, market studies, process evaluations, and market analyses with an evaluation to help guide portfolio investment decisions on what types of efforts to continue or fund in the future. As part of this, a gap analysis as a formative portfolio evaluation could be conducted to inform what types of programs, sectors, delivery mechanisms, or technologies to request for the next funding cycle.

Best Practice Studies

Another type of overarching study is the assessment of “best practices” within the energy program design and implementation arena. These studies document, for example, the best technology specific or customer type specific delivery approaches, such as the best lighting delivery approach for the small commercial sector, or the best marketing or customer contact approach, or the best measurement and evaluation techniques for certain types of technologies, programs or markets. These reports not only identify the key lessons learned in designing and implementing energy programs, but also provide guidelines for using the best practices. These studies help speed the rate of diffusion of the best practice in energy program design by identifying “best practices” and informing others about the practices and the results from the use of the practice.

Market Analyses

Market Structure and Operations Studies

Energy programs are typically designed for small sections of California rather than an area that is normally considered a technology marketing area. For example, the California PGC funded programs have significantly more programs that target a small section of the state (utility market area, bay area, hard-to-reach area, language-targeted neighborhoods, Orange County, etc.) than programs that are offered statewide. However, few providers of energy consuming equipment set market boundaries for their corporate marketing efforts by these types of boundaries. Rather, most technology manufacturing and distribution companies organize their market areas around considerations for total cost, available profit and status of the competition within an associated distribution network. When energy programs attempt to affect a market by changing the way product choice and use decisions are made, most programs are essentially trying to change the behaviors within a section of a larger market.

If portfolio designers and policy makers are to understand the ability of energy programs to affect a market, they need to know about the characteristics of the markets they are trying to effect. Market structure and operation studies might be needed, for example, to help plan what types of programs are needed in particular markets. These studies typically go beyond single program-based market assessment initiatives and focus on the structure, function and operations of a larger market in order to help assess the ability of a set of programs to influence a smaller portion of a larger market. This is especially

important for market transformation efforts that are trying to change the normal operations of a market to be more energy efficient. Individual programs that focus on small sections of a much larger market will not have the same potential as a portfolio of programs targeting that market or a larger market. Without information on how markets are structured and operate, it is difficult to know if a portfolio of programs has the capability to make a lasting change in that market.

As a result, there is a need to conduct market assessments to support a portfolio decision in addition to market assessments that an individual program might need to design their program, or that an individual program evaluation might conduct to identify the effects of an individual program on a given market. The portfolio manager may decide that certain types of market analyses should be conducted at the portfolio level.

Market Share Tracking Studies

In addition to market structure and operations studies there is also a need to conduct overarching studies that track market shares of energy efficient technologies. These studies allow policy makers and portfolio managers to assess the ability of a set of programs to achieve additional market share. If the market share for a given energy efficiency technology is already high, then additional program efforts may not have a significant effect on influencing changes to that share. Likewise, if the market share for an energy efficient technology is low, and there is evidence that energy programs can be designed to increase penetration cost-effectively, such that policy makers may want to encourage programs to address this area. These studies use technology tracking from a portfolio perspective in that they examine market share as well as the consequences associated with market changes and program interventions. This data can allow additional portfolio analyses to assess where programs may be beneficial to push a specific technology in order to acquire cost-effective resources.

Updating of Key Parameters

In the design and evaluation of programs, energy savings and cost-effectiveness calculations are based on key parameters that need to be updated periodically. These key parameters are the type of data that cut across several programs addressing multiple markets with multiple technologies rather than parameters that only apply to one or two programs. These studies would be better and more cost-efficiently addressed using overarching evaluation studies rather than a set of program-specific evaluations. For example, measure costs and cost differentials, avoided costs, load shape and peak demand information can be used with data from energy savings impact evaluations to estimate and understand a wider range of effects on participants. Elements within the Database for Energy Efficiency Resources (DEER)⁴⁰⁴ are critical portfolio planning parameters that need to be reviewed and updated over time so that ex-ante energy savings projections can be more accurate. As energy price and supply conditions change, it is

⁴⁰⁴ (Xenergy, ADM Associates et al. 2001).

important to change the way avoided costs are calculated and the time periods for which these costs apply so that cost-effectiveness assessments can reflect current and expected time and load sensitive price conditions. Similarly, as energy technology prices change with the movement of the market there is a need to periodically identify the cost difference between the standard technology and the energy efficient product choices so that program incentives and program designs can be tailored to match market conditions. Likewise, as these overarching studies are conducted it is appropriate to update the deemed savings estimates and the DEER to allow the use of better energy savings estimates consistent with each technology.

Evaluation Methodology Development

As in any research field, evaluation methodologies need to be periodically reviewed and updated due to advancements in evaluation theory, practice, and measurement technologies. A review of evaluation approaches and methods and applicability to specific kinds of programs and research goals should be conducted periodically as the evaluation field evolves. This Framework represents one of these types of studies that examines evaluation methods and approaches in order to help assure that the best evaluation approaches are used to understand the energy supply potential and implementation effects of California programs.

Examples of these types of studies that might be considered for the future include:

- A vintage study on impact evaluations to learn what circumstances determine how often an impact evaluation needs to be conducted;
- A benefit-cost analysis of standardization approaches and the development and use of Designated Unit of Measurement (DUM);
- A study to assess the best, least biased methods to econometrically derive net-to-gross estimates;
- A study to develop a standardized survey method and analysis approach for establishing net-to-gross estimates, perhaps calibrated to one or more econometric approaches for use by smaller programs; and
- Studies to assess and support better linkages in the translation of evaluation outputs into procurement/supply analysis inputs.

Managing and Funding Overarching Studies

Overarching studies are different than program-specific evaluations and need to be budgeted and implemented outside of the program-specific evaluation framework. Because these studies complement the ability of the energy efficiency or energy procurement statewide portfolio of programs, they need to be organized from this higher-level portfolio or statewide energy supply and acquisition perspective. These studies

need to be planned and directed by an organization that has an overview perspective of the programs, technologies and services provided statewide, and an understanding of the potential effects of a wide range of programs. For these reasons, an organization like the CPUC or CALMAC can be considered for managing and funding overarching studies that benefit the ability of energy programs to achieve statewide energy resources.

Appendix A: Glossary of Terms

ACCESS CHARGE - A charge paid by all market participants withdrawing energy from the ISO controlled grid. The access charge will recover the portion of a utility's transmission revenue requirement not recovered through the variable usage charge.

ACCURACY - An indication of how close some value is to the true value of the quantity in question. The term could also be used in reference to a model, or a set of measured data, or to describe a measuring instrument's capability.

ACHIEVABLE POTENTIAL - The amount of savings that can occur in response to specific program designs and delivery approaches, including program funding and measure incentive levels. Achievable potential studies are sometimes referred to as Market Potential studies.

ADDITION - An alteration to an existing building that increases conditioned space.

ADMINISTRATOR - A person, company, partnership, corporation, association, or other entity selected by the Commission and any Subcontractor that is retained by an aforesaid entity to contract for and administer energy efficiency programs funded in whole or in part from electric or gas Public Goods Charge (PGC) funds. For purposes of implementing PU Code Section 381.1, an "administrator" is any party that receives funding for and implements energy efficiency programs pursuant to PU Code Section 381. Similarly, a person, company, or other entity selected to contract and administer energy efficiency programs funded by procurement funds.

AFTER MARKET - Broad term that applies to any change after the original purchase, such as adding equipment not a part of the original purchase. As applied to alternative fuel vehicles, it refers to conversion devices or kits for conventional fuel vehicles.

AFUE – See ANNUAL FUEL UTILIZATION EFFICIENCY.

AIR CHANGE - The replacement of a quantity of air in a space within a given period of time, typically expressed as air changes per hour. If a building has one air change per hour, this is equivalent to all of the air in the building being replaced in a one-hour period.

AIR CONDITIONING COMFORT - Treating air to control its temperature, relative humidity, cleanliness, and distribution to meet the comfort requirements of the occupants of the conditioned space. Some air conditioners may not include all of these controls.

ALTERNATIVE ENERGY SOURCES - See RENEWABLE ENERGY.

AMBIENT AIR TEMPERATURE - Surrounding temperature, such as the outdoor air temperature around a building.

AMPERE (Amp) - The unit of measure that tells how much electricity flows through a conductor. It is like using cubic feet per second to measure the flow of water. For example, a 1,200 watt, 120-volt hair dryer pulls 10 amperes of electric current (watts divided by volts).

ANALYSIS OF COVARIANCE (ANCOVA) MODELS - A type of regression model also referred to as a “fixed effects” model. This model allows each individual to act as its own control. The unique effect of the stable, but unmeasured characteristics of each customer are their “fixed effects” from which this method takes its name. These fixed effects are held constant.

ANNUAL FUEL UTILIZATION EFFICIENCY (AFUE) - A measure of heating efficiency, in consistent units, determined by applying the federal test method for furnaces. This value is intended to represent the ratio of heat transferred to the conditioned space by the fuel energy supplied over one year. (See California Code of Regulations, Title 20, Section 1602(d)(1).)

ANNUAL MAXIMUM DEMAND - The greatest of all demands of the electrical load which occurred during a prescribed interval in a calendar year.

ANSI - American National Standards Institute is the national organization that coordinates development and maintenance of consensus standards and sets rules for fairness in their development. ANSI also represents the USA in developing international standards.

APPLIANCE EFFICIENCY STANDARDS - California Code of Regulations, Title 20, Chapter 2, Subchapter 4: Energy Conservation, Article 4: Appliance Efficiency Standards. Appliance Efficiency Standards regulate the minimum performance requirements for appliances sold in California and apply to refrigerators, freezers, room air conditioners, central air conditioners, gas space heaters, water heaters, plumbing fittings, fluorescent lamp ballasts and luminaires, and ignition devices for gas cooking appliances and gas pool heaters. New National Appliance Standards are in place for some of these appliances and will become effective for others at a future date.

APPLIANCE SATURATION - A percentage telling what proportion of all households in a given geographical area have a certain appliance.

AREA LOAD - The total amount of electricity being used at a given point in time by all consumers in a utility’s service territory.

ASHRAE - Acronym for American Society of Heating, Refrigerating and Air-Conditioning Engineers.

AUTOCORRELATION - The breakdown in the assumptions that the errors in regression analysis are uncorrelated due to correlation in the error term across observations in a time-series or cross-series, the error in one time period is directly correlated to the error

in another time period or cross-sectional category. First-order serial correlation is where that correlation is with the error in the subsequent/preceding time period. The correlation can be positive or negative.

AVERAGE COST - The revenue requirement of a utility divided by the utility's sales. Average cost typically includes the costs of existing power plants, transmission, and distribution lines, and other facilities used by a utility to serve its customers. It also includes operating and maintenance, tax, and fuel expenses.

AVERAGE DEMAND - The energy demand in a given geographical area over a period of time. For example, the number of kilowatt-hours used in a 24-hour period, divided by 24, tells the average demand for that period.

AVOIDED COST - (Regulatory) The amount of money that an electric utility would need to spend for the next increment of electric generation to produce or purchase elsewhere the power that it instead buys from a cogenerator or small-power producer. Federal law establishes broad guidelines for determining how much a qualifying facility (QF) gets paid for power sold to the utility.

BASE LOAD - The lowest level of power production needs during a season or year.

BASE RATE - That portion of the total electric or gas rate covering the general costs of doing business unrelated to fuel expenses.

BASELINE DATA - The measurements and facts describing facility operations and design during the baseline period. This will include energy use or demand and parameters of facility operation that govern energy use or demand.

BASELINE FORECAST - A prediction of future energy needs which does not take into account the likely effects of new conservation programs that have not yet been started.

BASELINE MODEL - The set of arithmetic factors, equations, or data used to describe the relationship between energy use or demand and other baseline data. A model may also be a simulation process involving a specified simulation engine and set of input data.

BASELINE PERIOD - The period of time selected as representative of facility operations before retrofit.

BENCHMARKING - A process that compares a 'set of results' against industry best practices.

BIENNIAL REPORT - The report issued by the California Energy Commission to the Governor and the Legislature every odd-numbered year assessing California's energy industry. The Biennial Report is supported by four policy documents that are issued every even-numbered year: the Electricity Report, the Fuels Report, the Conservation (or Efficiency) Report, and the Energy Development Report.

BILATERAL CONTRACT - A two-party agreement for the purchase and the sale of energy products and services.

BILLING DATA - Has multiple meanings. Metered data obtained from the electric or gas meter used to bill the customer for energy used in a particular billing period. Such meters typically conform to regulatory standards established for each customer class. Also used to describe the data representing the bills customers receive from the energy provider and also used to describe the customer billing and payment streams associated with customer accounts. This term is used to describe both consumption, demand, and account billing and payment information.

BILLING DEMAND - The demand used to calculate the demand charge cost. This is very often the monthly peak demand of the customer, but it may have a floor of some percentage of the highest monthly peak of the previous several months (a demand “ratchet”). May have other meanings associated with customer account billing practices.

BIOMASS - Energy resources derived from organic matter. These include wood, agricultural waste and other living-cell material that can be burned to produce heat energy. They also include algae, sewage and other organic substances that may be used to make energy through chemical processes.

BRITISH THERMAL UNIT (Btu) - The standard measure of heat energy. It takes one Btu to raise the temperature of one pound of water by one degree Fahrenheit at sea level. For example, it takes about 1,000 BTUs to make a pot of coffee. One Btu is equivalent to 252 calories, 778 foot-pounds, 1055 joules, and 0.293 watt-hours. Note: the abbreviation is seen as “Btu” or “BTU” interchangeably.

BROADCAST MESSAGE - A message (typically an ad) broadcast over a mass medium such as television, radio or newsprint.

BUILDING COMMISSIONING - Building commissioning provides documented confirmation that building systems as constructed function in accordance with the intent of the building designers, and satisfy the owner’s operational needs.

BUILDING ENERGY EFFICIENCY STANDARDS - California Code of Regulations (California Code of Regulations), Title 24, Part 2, Chapter 2-53; regulating the energy efficiency of buildings constructed in California.

BUILDING ENERGY SIMULATION MODEL - Computer models based on physical engineering principals and/or standards used to estimate energy usage and/or savings. These models do not make use of billing or metered data, but usually incorporate site-specific data on customers and physical systems. Building Simulation Models usually require such site-specific data as square footage, weather, surface orientations, elevations, space volumes, construction materials, equipment use, lighting, and building occupancy. Building simulation models can usually account for interactive effects between end uses (e.g., lighting and HVAC), part-load efficiencies, and changes in external and internal

heat gains/losses. Examples of building simulation models include ADM2, BLAST, and DOE-2.

BUILDING ENVELOPE - The assembly of exterior partitions of a building which enclose conditioned spaces, through which thermal energy may be transferred to or from the exterior, unconditioned spaces, or the ground. (See California Code of Regulations, Title 24, Section 2-5302.)

CADMAC - See CALIFORNIA DEMAND-SIDE MANAGEMENT MEASUREMENT ADVISORY COUNCIL.

CALIFORNIA CONSUMER POWER AND CONSERVATION FINANCING AUTHORITY (CPA) - The state agency charged with the responsibility to ensure sufficient electricity at reasonable market prices.

CALIFORNIA MEASUREMENT ADVISORY COUNCIL (CALMAC) - An informal committee made up of utility representatives, the California Public Utilities Commission, the California Energy Commission, and the Natural Resources Defense Council. CALMAC provides a forum for the development, implementation, presentation, discussion, and review of regional and statewide market assessment and evaluation studies for California energy efficiency programs conducted by member organizations using Public Goods Charge funds.

CALIFORNIA DEMAND-SIDE MANAGEMENT MEASUREMENT ADVISORY COUNCIL (CADMAC) - An informal committee made up of utility representatives, the Office of Ratepayer Advocates and the California Energy Commission. The purpose of the committee is to: provide a forum for presentations, discussions, and review of Demand Side Management (DSM) program measurement studies underway or completed; to coordinate the development and implementation of measurement studies common to all or most of the utilities; and to facilitate the development of effective, state-of-the-art protocols for measuring and evaluating the impacts of DSM programs.

CALIFORNIA ENERGY COMMISSION (CEC) - The state agency established by the Warren-Alquist State Energy Resources Conservation and Development Act in 1974 (Public Resources Code, Sections 25000 et seq.) responsible for energy policy. Funding for the Commission's activities comes from the Energy Resources Program Account, Federal Petroleum Violation Escrow Account and other sources. The CEC has statewide power plant siting, supply and demand forecasting, as well as multiple types of energy policy and analysis responsibilities.

CALIFORNIA PUBLIC UTILITIES COMMISSION (CPUC) - A state agency created by constitutional amendment in 1911 to regulate the rates and services of more than 1,500 privately owned utilities and 20,000 transportation companies. The CPUC is an administrative agency that exercises both legislative and judicial powers; its decisions and orders may be appealed only to the California Supreme Court. The major duties of the CPUC are to regulate privately owned utilities, securing adequate service to the

public at rates that are just and reasonable both to customers and shareholders of the utilities; including rates, electricity transmission lines and natural gas pipelines. The CPUC also provides electricity and natural gas forecasting, and analysis and planning of energy supply and resources. Its main headquarters are in San Francisco.

CALIFORNIA UTILITY RESEARCH COUNCIL (CURC) - Public Utilities Code, Sections 9201-9203 requires the California Energy Commission, the California Public Utilities Commission, and the investor-owned utilities (Pacific Gas and Electric Company, Southern California Edison, and San Diego Gas & Electric) to coordinate and promote consistency of research, development and demonstration (RD&D) programs with state energy policy. The CURC provides coordination for and sharing of information on energy RD&D in California to avoid duplication of efforts.

CALMAC – See CALIFORNIA MANAGEMENT MEASUREMENT ADVISORY COUNCIL.

CAPACITY - The amount of electric power for which a generating unit, generating station, or other electrical apparatus is rated either by the user or manufacturer. The term is also used for the total volume of natural gas that can flow through a pipeline over a given amount of time, considering such factors as compression and pipeline size.

CAPACITY FACTOR - A percentage that tells how much of a power plant's capacity is used over time. For example, typical plant capacity factors range as high as 80 percent for geothermal and 70 percent for cogeneration.

CAULKING - Material used to make an airtight seal by filling in cracks, such as those around windows and doors.

CCR - California Code of Regulations.

CEC - See CALIFORNIA ENERGY COMMISSION.

CFCs (CHLOROFLUOROCARBONS or CHLORINATED FLUOROCARBONS) - A family of artificially produced chemicals receiving much attention for their role in stratospheric ozone depletion. On a per molecule basis, these chemicals are several thousand times more effective as greenhouse gases than carbon dioxide. Since they were introduced in the mid-1930s, CFCs have been used as refrigerants, solvents and in the production of foam material.

CFM (cubic feet per minute) - A measure of flow rate.

CHANGE MODEL - A type of billing analysis designed to explain changes in energy usage. This can take the form of having the change in energy consumption (pre versus post) as the dependent variable (e.g., December pre-retrofit usage – December post-retrofit usage), or having consumption as the dependent variable and pre-retrofit consumption as one of the independent variables.

COEFFICIENT OF PERFORMANCE (COP) COOLING - The ratio of the rate of heat removal to the rate of energy input in consistent units, for a complete cooling system or factory assembled equipment, as tested under a nationally recognized standard or designated operating conditions.

COEFFICIENT OF PERFORMANCE (COP) HEATING & HEAT PUMPS - The ratio of the rate of heat delivered to the rate of energy input, in consistent units, for a complete heat pump system under designated operating conditions. Supplemental heat shall not be considered when checking compliance with the heat pump equipment COPs.

COINCIDENT DEMAND - The metered demand of a device, circuit, or building that occurs at the same time as the peak demand of the building or facility or at the same time as some other peak of interest, such as a utility's system load. This should properly be expressed so as to indicate the peak of interest, e.g., "demand coincident with the building peak."

COMFORT CONDITIONING - The process of treating air to simultaneously control its temperature, humidity, cleanliness, and distribution to meet the comfort requirements of the occupants of the conditioned space.

COMFORT ZONE - The range of temperatures over which the majority of persons feel comfortable (neither too hot nor too cold).

COMMERCIALIZATION - Programs or activities that increase the value or decrease the cost of integrating new products or services into the electricity sector.

COMPARISON GROUP - A group of customers who did not participate during the program year and who share as many characteristics as possible with the participant group.

COMPREHENSIVE - A program or project designed to achieve all cost-effective energy efficiency activities in individual buildings, usually including multiple energy efficiency measures.

CONDITIONAL DEMAND ANALYSIS (CDA) - A type of billing analysis in which observed energy consumption is estimated as a function of major end uses, often portrayed as dummy variables for their existence at the customer residence/facility.

CONDITIONED FLOOR AREA - The floor area of enclosed conditioned spaces on all floors measured from the interior surfaces of exterior partitions for non-residential buildings and from the exterior surfaces of exterior partitions for residential buildings. (See California Code of Regulations, Title 24, Section 2-5302.)

CONDITIONED SPACE - Enclosed space that is either directly conditioned space or indirectly conditioned space. (See California Code of Regulations, Title 24, Section 2-5302.)

CONDITIONED SPACE, DIRECTLY - An enclosed space that is provided with heating equipment that has a capacity exceeding 10 Btus/(hr-ft²), or with cooling equipment that has a capacity exceeding 10 Btus/(hr-ft²). An exception is if the heating and cooling equipment is designed and thermostatically controlled to maintain a process environment temperature less than 65 degrees Fahrenheit or greater than 85 degrees Fahrenheit for the whole space the equipment serves. (See California Code of Regulations, Title 24, Section 2- 5302.)

CONDITIONED SPACE, INDIRECTLY - Enclosed space that: (1) has a greater area weighted heat transfer coefficient (u-value) between it and directly conditioned spaces than between it and the outdoors or unconditioned space; (2) has air transferred from directly conditioned space moving through it at a rate exceeding three air changes per hour.

CONSERVATION - Steps taken to cause less energy to be used than would otherwise be the case. These steps may involve improved efficiency, avoidance of waste, reduced consumption, etc. They may involve installing equipment (such as a computer to ensure efficient energy use), modifying equipment (such as making a boiler more efficient), adding insulation, changing behavior patterns, etc.

CONSTRUCT VALIDITY - The extent to which an operating variable/instrument accurately taps an underlying concept/hypothesis, properly measuring an abstract quality or idea.

CONTENT VALIDITY - The extent to which an operating measure taps all the separate sub-concepts of a complicated concept.

CONTINGENCY PLANNING - The Energy Commission's strategy to respond to impending energy emergencies such as curtailment or shortage of fuel or power because of natural disasters or the result of human or political causes, or a clear threat to public health, safety or welfare. The contingency plan specifies state actions to alleviate the impacts of a possible shortage or disruption of petroleum, natural gas or electricity. The plan is reviewed and updated at least every five years, with the last plan being adopted in 1993. Legislative authority for the California Energy Shortage Contingency Plan is found in Public Resources Code, Section 25216.5.

CONTRACT DEMAND - The maximum demand, which may or may not be metered, that is expected or allowed under the contract with the utility providing the energy.

CONVERGENT VALIDITY - When two instruments/questions/measurement methods obtain similar results when measuring the same underlying construct with varying questions/approaches.

CONVERSION FUEL FACTOR - A number stating units of one system in corresponding values of another system.

COOLING CAPACITY, TOTAL - Available refrigerating capacity of an air conditioner for removing sensible heat and latent heat from the space to be conditioned.

COOLING DEGREE DAYS - The cumulative number of degrees in a month or year by which the mean temperature is above 18.3°C/65°F.

COOLING LOAD - The rate at which heat must be extracted from a space in order to maintain the desired temperature within the space.

COOLING LOAD TEMPERATURE DIFFERENCE (CLTD) - A value used in cooling load calculations for the effective temperature difference (delta T) across a wall or ceiling, which accounts for the effect of radiant heat as well as the temperature difference.

COP - See **COEFFICIENT OF PERFORMANCE**.

CORRELATION COEFFICIENT - A measure of the linear association between two variables, calculated as the square root of the R^2 obtained by regressing one variable on the other and signed to indicate whether the relationship is positive or negative.

COST-EFFECTIVENESS - An indicator of the relative performance or economic attractiveness of any energy efficiency investment or practice when compared to the costs of energy produced and delivered in the absence of such an investment. In the energy efficiency field, the present value of the estimated benefits produced by an energy efficiency program as compared to the estimated total program's costs, from the perspective of either society as a whole or of individual customers, to determine if the proposed investment or measure is desirable from a variety of perspectives, e.g., whether the estimated benefits exceed the estimated costs. See **TOTAL RESOURCE COST TEST – SOCIETAL VERSION** and **PARTICIPANT COST TEST**.

CPA - See **CALIFORNIA CONSUMER POWER AND CONSERVATION FINANCING AUTHORITY**.

CPUC - See **CALIFORNIA PUBLIC UTILITIES COMMISSION**.

CREAM SKIMMING - Cream skimming results in the pursuit of only the lowest cost or most cost-effective energy efficiency measures, leaving behind other cost-effective opportunities. Cream skimming is inappropriate when lost opportunities are created in the process.

CROSS-CUTTING PROGRAM - A program that involves any or all of the following: multiple customer types (residential and/or non-residential), and/or multiple building types (retrofit, remodeling, and/or new construction).

CUBIC FOOT - The most common unit of measurement of natural gas volume. It equals the amount of gas required to fill a volume of one cubic foot under stated conditions of

temperature, pressure and water vapor. One cubic foot of natural gas has an energy content of approximately 1,000 Btus. One hundred (100) cubic feet equals one therm (100 ft³ = 1 therm).

CURC - See CALIFORNIA UTILITY RESEARCH COUNCIL.

CUSTOMER - Any person or entity that pays an electric and/or gas bill to an IOU and that is the ultimate consumer of goods and services including energy efficiency products, services, or practices.

CUSTOMER INFORMATION - Non-public information and data specific to a utility customer which the utility acquired or developed in the course of its provision of utility services.

DAYLIGHTING - The use of sunlight to supplement or replace electric lighting.

DEER - Database for Energy Efficient Resources.

DEFAULT ASSUMPTION - The value of an input used in a calculation procedure when a value is not entered by the designer.

DELTA - Difference in temperature. Often used in the context of the difference between the design indoor temperature and the outdoor temperature.

DEMAND - The time rate of energy flow. Demand usually refers to electric power and is measured in kW (equals kWh/h) but can also refer to natural gas, usually as Btu/hr, kBtu/hr, therms/day or ccf/day.

DEMAND (Utility) - The level at which electricity or natural gas is delivered to users at a given point in time. Electric demand is expressed in kilowatts.

DEMAND BILLING - The electric capacity requirement for which a large user pays. It may be based on the customer's peak demand during the contract year, on a previous maximum or on an agreed minimum. Measured in kilowatts.

DEMAND CHARGE - The sum to be paid by a large electricity consumer for its peak usage level.

DEMAND RESPONSIVENESS - Also sometimes referred to as load shifting. Activities or equipment that induce consumers to use energy at different (lower cost) times of day or to interrupt energy use for certain equipment temporarily, usually in direct response to a price signal. Examples: interruptible rates, doing laundry after 7 p.m., air conditioner recycling programs.

DEMAND SAVINGS - The reduction in the demand from the pre-retrofit baseline to the post-retrofit demand, once independent variables (such as weather or occupancy) have

been adjusted for. This term is usually applied to billing demand, to calculate cost savings, or to peak demand, for equipment sizing purposes.

DEMAND SIDE MANAGEMENT (DSM) - The methods used to manage energy demand including energy efficiency, load management, fuel substitution and load building. See **LOAD MANAGEMENT**.

DEMONSTRATION - The application and integration of a new product or service into an existing or new system. Most commonly, demonstration involves the construction and operation of a new electric technology interconnected with the electric utility system to demonstrate how it interacts with the system. This includes the impacts the technology may have on the system and the impacts that the larger utility system might have on the functioning of the technology.

DESIGN ASSISTANCE - These programs provide design and analysis services to the architects and engineers responsible for the design of new residential and commercial buildings. The goal of these programs is to make the building more energy efficient.

DESIGNATED UNIT(S) OF MEASUREMENT - The metric(s) used for expressing load impacts for a particular program as a function of customer characteristics (e.g., kWh per square foot). The metric is used to compare load impacts of different customers or customer groups (participants, comparison group samples, future participants).

DIRECT INSTALLATION PROGRAMS - These types of programs provide free energy efficiency measures for qualified customers. Typical measures distributed by these programs include low flow showerheads and compact fluorescent bulbs.

DIRECT SOLAR GAIN - Solar energy collected from the sun (as heat) in a building through windows, walls, skylights, etc.

DISTRIBUTED GENERATION - A distributed generation system involves small amounts of generation located on a utility's distribution system for the purpose of meeting local (substation level) peak loads and/or displacing the need to build additional (or upgrade) local distribution lines.

DOUBLE-BARRELED QUESTIONS - A poorly worded questionnaire item which actually asks two questions at the same time, thereby not allowing unique and accurate interpretation of the results.

DOUBLE GLAZING - Windows having two sheets of glass with an airspace between.

DOUBLE DIPPING - Taking advantage of multiple financial incentives offered by multiple programs for undertaking only one activity.

DRY BULB TEMPERATURE - A measure of the sensible temperature of air.

DSM - See DEMAND SIDE MANAGEMENT.

DUAL-PANED (double-glazed) - Two panes of glass or other transparent material, separated by a space.

ECONOMIC EFFICIENCY - A term that refers to the optimal production and consumption of goods and services. This generally occurs when prices of products and services reflect their marginal costs. Economic efficiency gains can be achieved through cost reduction, but it is better to think of the concept as actions that promote an increase in overall net value (which includes, but is not limited to, cost reductions).

ECONOMIC POTENTIAL - Refers to that part of the technical potential that can cost-effectively be obtained when compared to supply-side alternatives.

ECONOMIES OF SCALE - Economies of scale exist where the industry exhibits decreasing average long-run costs with size.

EDUCATION PROGRAMS - Programs primarily intended to educate customers about energy efficient technologies or behaviors or provide information about programs that offer energy efficiency or load reduction information or services.

EER - See ENERGY EFFICIENCY RATIO.

EFFECTIVE USEFUL LIFE (EUL) - An estimate of the median number of years that the measures installed under the program are still in place and operable.

EFFICACY, LIGHTING - The ratio of light from a lamp to the electrical power consumed, including ballast losses, expressed as lumens per watt. (See California Code of Regulations, Title 24, Section 2-5302.)

EFFICIENCY - The ratio of the useful energy delivered by a dynamic system (such as a machine, engine, or motor) to the energy supplied to it over the same period or cycle of operation. The ratio is usually determined under specific test conditions.

ELECTRIC PUBLIC GOODS CHARGE (PGC) - Per Assembly Bill (AB) 1890, a universal charge applied to each electric utility Customer's bill to support the provision of public goods. Public goods covered by California's electric PGC include public purpose energy efficiency programs, low-income services, renewables, and energy-related research and development.

EMISSIVITY - The property of emitting radiation; possessed by all materials to a varying extent.

EMITTANCE - The emissivity of a material, expressed as a fraction. Emittance values range from 0.05 for brightly polished metals to 0.96 for flat black paint.

END USE (MEASURES/GROUPS) - Refers to a broad or sometimes narrower category that the program is concentrating efforts upon. Examples of end uses include: refrigeration, food service, HVAC, appliances, envelope and lighting.

ENERGY CONSUMPTION - The amount of energy consumed in the form in which it is acquired by the user. The term excludes electrical generation and distribution losses.

ENERGY COST - The total cost for energy, including such charges as base charges, demand charges, customer charges, power factor charges, and miscellaneous charges.

ENERGY (FUEL) DIVERSITY - Policy that encourages the development of energy technologies to diversify energy supply sources, thus reducing reliance on conventional (petroleum) fuels; applies to all energy sectors.

ENERGY EFFICIENCY - Using less energy/electricity to perform the same function. Programs designed to use electricity more efficiently - doing the same with less. For the purpose of this paper, energy efficiency is distinguished from DSM programs in that the latter are utility sponsored and financed, while the former is a broader term not limited to any particular sponsor or funding source. "Energy conservation" is a term that has also been used but it has the connotation of doing without in order to save energy rather than using less energy to do the same thing and so is not used as much today. Many people use these terms interchangeably.

ENERGY EFFICIENCY IMPROVEMENT - Reduced energy use for a comparable level of service, resulting from the installation of an energy efficiency measure or the adoption of an energy efficiency practice. Level of service may be expressed in such ways as the volume of a refrigerator, temperature levels, production output of a manufacturing facility, or lighting level per square foot.

ENERGY EFFICIENCY MEASURE - Installation of equipment, subsystems, or systems, or modification of equipment, subsystems, systems, or operations, on the customer side of the meter, for the purpose of reducing energy and/or demand (and, hence, energy and/or demand costs) at a comparable level of service.

ENERGY EFFICIENCY OF A MEASURE - A measure of the energy used to provide a specific service or to accomplish a specific amount of work (e.g., kWh per cubic foot of a refrigerator, therms per gallon of hot water).

ENERGY EFFICIENCY OF EQUIPMENT - The percentage of gross energy input that is realized as useful energy output of a piece of equipment.

ENERGY EFFICIENCY PRACTICE - The use of high-efficiency products, services, and practices or an energy-using appliance or piece of equipment, to reduce energy usage while maintaining a comparable level of service when installed or applied on the customer side of the meter. Energy efficiency activities typically require permanent

replacement of energy-using equipment with more efficient models. Examples: refrigerator replacement, light fixture replacement, cooling equipment upgrades.

ENERGY EFFICIENCY RATIO (EER) - The ratio of cooling capacity of an air conditioning unit in Btus per hour to the total electrical input in watts under specified test conditions. (See California Code of Regulations, Title 20, Section 1602(c)(6).)

ENERGY MANAGEMENT SERVICES - Programs intended to provide customer assistance in the form of information on the relative costs and benefits to the customer of installing measures or adopting practices which can reduce the customer's utility bills. The information is solicited by the customer and recommendations are based on the customer's recent billing history and/or customer-specific information regarding appliance and building characteristics.

ENERGY MANAGEMENT SYSTEM - A control system (often computerized) designed to regulate the energy consumption of a building by controlling the operation of energy consuming systems, such as the heating, ventilation and air conditioning (HVAC), lighting, and water heating systems.

ENERGY RESOURCES PROGRAM ACCOUNT (ERPA) - The state law that directs California electric utility companies to gather a state energy surcharge per kilowatt hour of electricity consumed by a customer. These funds are used for operation of the California Energy Commission. As of January 1, 2004, the surcharge is set at of three-tenths of one mil (\$0.0003) per kilowatt-hour.

ENERGY SAVINGS - The reduction in use of energy from the pre-retrofit baseline to the post-retrofit energy use, once independent variables (such as weather or occupancy) have been adjusted for.

ENERGY SERVICE COMPANY (ESCO) - An organization that designs, procures, installs, and possibly maintains one or more energy conservation measures (ECMs) at an owner's facility or facilities. Typically ESCOs offer to reduce a client's electricity consumption with the cost savings being split with the client.

ENGINEERING APPROACHES - Methods using engineering algorithms or models to estimate energy and/or demand use.

ENGINEERING USEFUL LIFE - An engineering estimate of the number of years that a piece of equipment will operate if properly maintained.

ERPA - See ENERGY RESOURCES PROGRAM ACCOUNT.

ERROR - Deviation of measurements from the true value.

ESCO - See ENERGY SERVICE COMPANY.

EUL - See EFFECTIVE USEFUL LIFE.

EVALUATION - The performance of studies and activities aimed at determining the effects of a program, or any of a wide range of assessment activities associated with understanding or documenting program performance or potential performance, assessing program or program-related markets and market operations, or any of a wide range of evaluative efforts including assessing program-induced changes in energy efficiency markets, levels of demand or energy savings, and program cost-effectiveness.

EXCEPTIONAL METHOD - An approved alternative calculation method that analyzes designs, materials, or devices that cannot be adequately modeled using public domain computer programs. Exceptional methods must be submitted to and approved by the California Energy Commission. [See California Code of Regulations, Title 20, Section 1409(b)3] Two examples of exceptional methods are the controlled ventilation crawl space (CVC) credit and the combined hydronic space and water heating method.

EXTERNAL VALIDITY - The extent to which the association between an independent variable and a dependent variable that is demonstrated within a research setting also holds true in the general environment.

FINANCING PROGRAMS - These types of programs encourage investments in energy efficiency through offerings such as below market interest rates or terms, gap financing, and financial packages.

FOOTCANDLE - A unit of illuminance on a surface that is one foot from a uniform point source of light of one candle and is equal to one lumen per square foot.

FRAMING EFFECTS - The effect of framing (wood or metal studs, joists, beams, etc.) on the overall U-value of a wall, roof, floor, window or other building surface. Framing generally increases the U-Value and decreases the R-Value of insulated surfaces.

FREE DRIVER - A non-participant who adopted a particular efficiency measure or practice as a result of a utility program. See SPILLOVER EFFECTS for aggregate impacts.

FREE RIDER - A program participant who would have implemented the program measure or practice in the absence of the program.

GAS PUBLIC GOODS CHARGE - Created by AB1002 in 2000, an unbundled rate component included on gas customer bills to fund public purpose programs including energy efficiency, low-income, and research and development.

GENERAL LIGHTING - Lighting designed to provide a substantially uniform level of illumination throughout an area, exclusive of any provision for special visual tasks or decorative effects. (See California Code of Regulations, Title 24, Section 2-5302.)

GIGAWATT (GW) - One thousand megawatts (1,000 MW), one million kilowatts (1,000,000 kW), or one billion watts (1,000,000,000 watts) of electricity. One gigawatt is enough to supply the electric demand of about one million average California homes.

GIGAWATT-HOUR (GWH) - One million kilowatt-hours of electric power.

GLAZING - A covering of transparent or translucent material (typically glass or plastic) used for admitting light.

GLOBAL CLIMATE CHANGE - Gradual changing of global climates due to buildup of carbon dioxide and other greenhouse gases in the earth's atmosphere. Carbon dioxide produced by burning fossil fuels has reached levels greater than what can be absorbed by green plants and the seas.

GREENHOUSE EFFECT - The presence of trace atmospheric gases make the earth warmer than would direct sunlight alone. These gases (carbon dioxide [CO₂], methane [CH₄], nitrous oxide [N₂O], tropospheric ozone [O₃], and water vapor [H₂O]) allow visible light and ultraviolet light (shortwave radiation) to pass through the atmosphere and heat the earth's surface. This heat is re-radiated from the earth in form of infrared energy (long wave radiation). The greenhouse gases absorb part of that energy before it escapes into space. This process of trapping the long wave radiation is known as the greenhouse effect. Scientists estimate that without the greenhouse effect, the earth's surface would be roughly 54 degrees Fahrenheit colder than it is today - too cold to support life as we know it. See GLOBAL CLIMATE CHANGE.

GROSS AREA - The area of a surface including areas not belonging to that surface (such as windows and doors in a wall).

GROSS LOAD IMPACT - The change in energy consumption and/or demand that results directly from program-related actions taken by participants in the DSM program, regardless of why they participated.

HARDWARE PROGRAMS - Programs primarily intended to provide measurable energy savings through installation of energy efficiency measures or provision of energy efficiency services.

HEAT BALANCE - The outdoor temperature at which a building's internal heat gain (from people, lights and machines) is equal to the heat loss through windows, roof and walls.

HEAT CAPACITY - The amount of heat necessary to raise the temperature of a given mass one degree. Heat capacity may be calculated by multiplying the mass by the specific heat.

HEAT GAIN - An increase in the amount of heat contained in a space, resulting from direct solar radiation, heat flow through walls, windows, and other building surfaces, and the heat given off by people, lights, equipment, and other sources.

HEAT LOSS - A decrease in the amount of heat contained in a space, resulting from heat flow through walls, windows, roof and other building surfaces and from exfiltration of warm air.

HEAT PUMP - An air conditioning unit which is capable of heating by refrigeration, transferring heat from one (often cooler) medium to another (often warmer) medium, and which may or may not include a capability for cooling. This reverse-cycle air conditioner usually provides cooling in summer and heating in winter.

HEAT RATE - A number that tells how efficient a fuel-burning power plant is. The heat rate equals the Btu content of the fuel input divided by the kilowatt-hours of power output.

HEAT TRANSFER - Flow of heat energy induced by a temperature difference. Heat flow through a building envelope typically flows from a heated or hot area, to a cooled or cold area.

HEATING DEGREE DAYS - The cumulative number of degrees in a month or year by which the mean temperature falls below 18.3°C/65°F.

HEAP - See HOME ENERGY ASSISTANCE PROGRAM.

HEATING SEASONAL PERFORMANCE FACTOR (HSPF) - A representation of the total heating output of a central air conditioning heat pump in Btus during its normal usage period for heating, divided by the total electrical energy input in watt-hours during the same period, as determined using the test procedure specified in the California Code of Regulations, Title 20, Section 1603(c).

HETEROSCEDASTICITY - The variance in the error term is changing. This violates the regression assumption of constant variance. A common example is where variance is expected to be greater on a variable measurement for larger firms than for smaller firms.

HOME ENERGY ASSISTANCE PROGRAM (HEAP) - A centrally operated direct payment program that assists eligible households in offsetting the cost of heating and cooling their homes. Payments are generally made in the form of dual party warrants (checks) made payable to the applicant and their designated utility company. The program is administered by the California Department of Economic Opportunity using federal and state funds.

HOMOSCEDASTIC (HOMOSCEDASTICITY) - The error term has constant variance, an assumption of classical regression analysis.

HORSEPOWER (HP) - A unit for measuring the rate of doing work. One horsepower equals about three-fourths of a kilowatt (745.7 watts).

HSPF - See HEATING SEASONAL PERFORMANCE FACTOR.

HVAC (Heating Ventilation and Air Conditioning) - A system that provides heating, ventilation and/or cooling within or associated with a building.

HVAC SYSTEM - The equipment, distribution network, and terminals that provides either collectively or individually the processes of heating, ventilating, or air conditioning to a building.

JACK-KNIFE - A means of estimating a statistical/econometric estimator's variance by computing the variance of the estimates produced by that estimator omitting each of the observations in turn.

IMPACT EVALUATION - Used to measure the change in energy and/or demand usage (such kWh, kW and therms) attributed to energy efficiency and demand response programs.

IMPACT YEAR - Depending on the context, impact year means either (a) the twelve months subsequent to participation used to represent program costs or load impacts occurring in that year, or (b) any calendar year after the program year in which load impacts may occur.

IMPLEMENTATION THEORY - A theory describing how a program should be structured and implemented and the theoretical rationale supporting the reasons for the program structure and the implementation approach.

IMPLEMENTER - An entity or person selected and contracted with or qualified by a program administrator or by the Commission to receive PGC funds for providing products and services to customers.

INCENTIVES - Financial support (e.g., rebates, low-interest loans) to install energy efficiency measures. The incentives are solicited by the customer and based on the customer's billing history and/or customer-specific information.

INDEPENDENT VARIABLES - The factors that affect the energy and demand used in a building but cannot be controlled (e.g., weather or occupancy).

INDIGENOUS ENERGY RESOURCES - Power and heat derived from sources native to California. These include geothermal, hydro, biomass, solar and wind energy. The term usually is understood to include cogeneration facilities.

INFILTRATION - The uncontrolled inward leakage of air through cracks and gaps in the building envelope, especially around windows, doors and duct systems.

INFILTRATION BARRIER - A material placed on the outside or the inside of exterior wall framing to restrict inward air leakage, while permitting the outward escape of water vapor from the wall cavity. (See California Code of Regulations, Title 24, Section 2-5302.)

INFORMATION PROGRAMS - Programs primarily intended to provide customers with information regarding generic (not customer-specific) conservation and energy efficiency opportunities. For these programs, the information may be unsolicited by the customer. Programs that provide incentives in the form of unsolicited coupons for discount on low cost measures are also included.

INSULATION, THERMAL - A material having a relatively high resistance of heat flow and used principally to retard heat flow. See R-VALUE.

INTEGRATED PART-LOAD VALUE (IPLV) - A single number figure of merit based on part-load EER or COP expressing part-load efficiency for air conditioning and heat pump equipment on the basis of weighted operation at various load capacities for the equipment.

INTEGRATED RESOURCE PLANNING (IRP) - A public planning process and framework within which the costs and benefits of both demand- and supply-side resources are evaluated to develop the least-total-cost mix of utility resource options. In many states, IRP includes a means for considering environmental damages caused by electricity supply/transmission and identifying cost-effective energy efficiency and renewable energy alternatives. IRP has become a formal process prescribed by law in some states and under some provisions of the Clean Air Act amendments of 1992.

INTERNAL VALIDITY - The extent to which alternative explanations can be eliminated as causes for an observed association between independent and dependent variable(s) within a research setting/sample.

INTERRUPTIBLE SERVICE (Electric utility) - Electricity supplied under agreements that allow the supplier to curtail or stop service at times.

INVESTOR-OWNED UTILITIES (IOU) - A private company that provides a utility, such as water, natural gas or electricity, to a specific service area. The investor-owned utility is regulated by the California Public Utilities Commission.

JOULE - A unit of work or energy equal to the amount of work done when the point of application of force of 1 newton is displaced 1 meter in the direction of the force. It takes 1,055 joules to equal a British thermal unit. It takes about 1 million joules to make a pot of coffee.

kBtu - One-thousand (1,000) Btus.

KILOWATT (kW) - One thousand (1,000) watts. A unit of measure of the amount of electricity needed to operate given equipment. On a hot summer afternoon a typical home with central air conditioning and other equipment in use might have a demand of four kW each hour.

KILOWATT-HOUR (kWh) - The most commonly used unit of measure telling the amount of electricity consumed over time. It means one kilowatt of electricity supplied for one hour.

LATENT HEAT - A change in the heat content that occurs without a corresponding change in temperature, usually accompanied by a change of state (as from liquid to vapor during evaporation).

LATENT LOAD - The cooling load caused by moisture in the air.

LEVEL OF SERVICES - The utility received by a customer from energy-using equipment. Level of service may be expressed as the volume of a refrigerator, an indoor temperature level, the production output of a manufacturing facility, lighting levels per square foot, etc.

LIFE EXTENSION - A term used to describe capital expenses which reduce operating and maintenance costs associated with continued operation of electric utility boilers. Such boilers usually have a forty-year operating life under normal circumstances.

LIFE-CYCLE COST - Amount of money necessary to own, operate and maintain a building over its useful life.

LIFELINE RATES - Rates charged by a utility company for the low income, the disadvantaged and senior citizens. The rates provide a discount for minimum necessary utilities, such as electricity requirements of typically 300 to 400 kilowatt/hours per month.

LINEAR REGRESSION ESTIMATOR (FOR SAMPLING) - Used to increase precision in a sample design by using an auxiliary variable that is correlated with the desired outcome variable where the relationship line does not go through the origin (as would be the case for a ratio estimator). Either a uniform regression estimator across strata or separate regression estimators per strata may be selected as the best design based upon the whether there is a relationship between the auxiliary variable and the selection of the strata parameter.

LIRM - See **LOAD IMPACT REGRESSION MODEL**.

LOAD - An end use device or an end use customer that consumes power. The amount of electric power supplied to meet one or more end user's needs. Load should not be confused with demand, which is the measure of power that a load receives or requires.

LOAD DIVERSITY - The condition that exists when the peak demands of a variety of electric customers occur at different times. This is the objective of “load molding” strategies, ultimately curbing the total capacity requirements of a utility.

LOAD FACTOR - A percent telling the difference between the amount of electricity a consumer used during a given time span and the amount that would have been used if the usage had stayed at the consumer’s highest demand level during the whole time. The term also is used to mean the percentage of capacity of an energy facility - such as a power plant or gas pipeline - that is utilized in a given period of time.

LOAD IMPACT - Changes in electric energy use, electric peak demand, or natural gas use.

LOAD IMPACT REGRESSION MODEL (LIRM) - The most general definition of a LIRM is a statistical model that produces estimates of the load impacts of energy conservation programs. Depending on the particular approach and the statistical issues encountered, it may involve more than one regression model and technique: (1) The load impact estimation model typically is a linear or non-linear regression model that uses billing data that estimates gross and/or net load impacts. Data from program non-participants, in addition to participant data, can be used to derive net impacts directly or to affect other statistical control. (2) The participant/decision model typically is a discrete choice model used in conjunction with the load impact estimation model to isolate free ridership effects, generate self-selection correction terms, and/or net-to-gross ratios as needed. When this model is used to estimate a net-to-gross ratio, the resulting estimate is multiplied by an estimate of gross load impact to yield an estimate of net load impact.

LOAD MANAGEMENT - Steps taken to reduce power demand at peak load times or to shift some of it to off-peak times. This may be with reference to peak hours, peak days or peak seasons. The main thing affecting electric peaks is air conditioning usage, which is therefore a prime target for load management efforts. Load management may be pursued by persuading consumers to modify behavior or by using equipment that regulates some electric consumption.

LOAD PROGRAM - A program that provides services to customers in only one jurisdiction of the state (e.g., one county, city, or region). Local programs may be experimental and are designed to serve the needs of a particular geographic area.

LOAD SHAPE - The time-of-use pattern of customer or equipment energy use. This pattern can be over 24 hours or over a year (8760 hours).

LOAD SHAPE IMPACTS - Changes in load shape induced by a program.

LOADED QUESTION - A question whose wording encourages the respondent to provide a particular answer.

LOGIC MODEL - The graphical representation of the program theory showing the flow between activities, their outputs, and subsequent short-term, intermediate, and long-term outcomes. Often the logic model is displayed with these elements in boxes and the causal flow being shown by arrows from one to the others in the program logic. It can also be displayed as a table with the linear relationship presented by the rows in the table.

LOSSES (Electric utility) - Electric energy or capacity that is wasted in the normal operation of a power system. Some kilowatt-hours are lost in the form of waste heat in electrical apparatus such as substation conductors. **LINE LOSSES** are kilowatts or kilowatt-hours lost in transmission and distribution lines under certain conditions.

LOST OPPORTUNITIES - Energy efficiency measures that offer long-lived, cost-effective savings that are fleeting in nature. A lost opportunity occurs when a customer does not install an energy efficiency measure that is cost-effective at the time, but whose installation is unlikely to be cost-effective (or is less cost-effective) later.

LOW-E - A special coating that reduces the emissivity of a window assembly, thereby reducing the heat transfer through the assembly.

LUMEN - A measure of the amount of light available from a light source equivalent to the light emitted by one candle.

LUMENS/WATT - A measure of the efficacy of a light fixture; the number of lumens output per watt of power consumed.

LUMINAIRE - A complete lighting unit consisting of a lamp or lamps together with the parts designed to distribute the light, to position and protect the lamps and to connect the lamps to the power supply. (See California Code of Regulations, Section 2- 1602(h).)

LUX - A unit of illumination equal to the direct illumination on a surface that is everywhere one meter from a uniform point source of one candle; a unit of illumination that is equal to one lumen per square meter.

MAIN METER - The meter that measures the energy used for the whole facility. There is at least one meter for each energy source and possibly more than one per source for large facilities. Typically, utility meters are used, but dataloggers may also be used as long as they isolate the load for the facility being studied. When more than one meter per energy source exists for a facility, the main meter may be considered the accumulation of all the meters involved.

MARGINAL COST - The sum that has to be paid for the next increment of product or service. The marginal cost of electricity is the price to be paid for kilowatt-hours above and beyond those supplied by presently available generating capacity.

MARKET ACTORS - Individuals and organizations in the production, distribution, and/or delivery chain of energy efficiency products, services and practices. This may

include, but is not limited to, manufacturers, distributors, wholesalers, retailers, vendors, dealers, contractors, developers, builders, financial institutions, and real estate brokers and agents.

MARKET ASSESSMENT - An analysis function that provides an assessment of how and how well a specific market or market segment is functioning with respect to the definition of well-functioning markets or with respect to other specific policy objectives. Generally includes a characterization or description of the specific market or market segments, including a description of the types and number of buyers and sellers in the market, the key actors that influence the market, the type and number of transactions that occur on an annual basis, and the extent to which energy efficiency is considered an important part of these transactions by market participants. This analysis may also include an assessment of whether or not a market has been sufficiently transformed to justify a reduction or elimination of specific program interventions. Market assessment can be blended with strategic planning analysis to produce recommended program designs or budgets. One particular kind of market assessment effort is a baseline study, or the characterization of a market before the commencement of a specific intervention in the market, for the purpose of guiding the intervention and/or assessing its effectiveness later.

MARKET BARRIER - Any characteristic of the market for an energy-related product, service, or practice that helps to explain the gap between the actual level of investment in, or practice of, energy efficiency and an increased level that would appear to be cost-beneficial to the consumer.

MARKET EFFECT - A change in the structure or functioning of a market or the behavior of participants in a market that result from one or more program efforts. Typically these efforts are designed to increase in the adoption of energy efficient products, services, or practices and are causally related to market interventions.

MARKET EVENT - The broader circumstances under which a customer considers adopting an energy efficiency product, service, or practice. Types of market events include, but are not necessarily limited to, the following: (a) new construction, or the construction of a new building or facility; (b) renovation, or the updating of an existing building or facility; (c) remodeling, or a change in an existing building; (d) replacement, or the replacement of equipment, either as a result of an emergency such as equipment failure, or as part of a broader planned event; and, (e) retrofit, or the early replacement of equipment or refitting of a building or facility while equipment is still functioning, often as a result of an intervention into energy efficiency markets.

MARKET PARTICIPANTS - The individuals and organizations participating in transactions with one another within an energy efficiency market or markets, including customers and market actors.

MARKET POTENTIAL STUDIES - See **ACHIEVABLE POTENTIAL**.

MARKET SECTORS - General types of markets that a program may target or in which a service offering may be placed. Market sectors include categories such as Agricultural, Commercial, Industrial, Government, and Institutional. Market sectors help the Commission assess how well its portfolio of programs is addressing the variety of markets for energy efficiency products and services in the state.

MARKET SEGMENTS - A part of a market sector that can be grouped together as a result of a characteristic similar to the group. Within the residential sector are market segments, such as renters, owners, multi-family, single-family, etc. These market segments help the Commission assess how well its portfolio of programs is addressing the variety of segments within the markets served.

MARKET THEORY - A theoretical description of how a market operates relative to a specific program or set of programs designed to influence that market. Market theories typically include the identification of key market actors, information flows, and product flows through the market, relative to a program designed to change the way the market operates. Market theories are typically grounded upon the information provided from a market assessment but can also be based on other information. Market theories often describe how a program intervention can take advantage of the structure and function of a market to transform the market. Market theories can also describe the key barriers and benefits associated with a market and describe how a program can exploit the benefits and overcome the barriers.

MARKET TRANSFORMATION - A reduction in market barriers resulting from a market intervention, as evidenced by a set of market effects, that lasts after the intervention has been withdrawn, reduced or changed.

MARKET-BASED PRICE - A price set by the mutual decisions of many buyers and sellers in a competitive market.

MCF - One thousand cubic feet of natural gas, having an energy value of one million Btu. A typical home might use six MCF in a month.

MEASURE (noun) - A product whose installation and operation at a customer's premises results in a reduction in the customer's on-site energy use, compared to what would have happened otherwise. See also **ENERGY EFFICIENCY MEASURE**.

MEASURE (verb) - Use of an instrument to assess a physical quantity, or use of a computer simulation to estimate a physical quantity.

MEASURE DATA - Data collected from participants in a utility efficiency program after their participation.

MEASURE RETENTION STUDY - An assessment of (a) the length of time the measure(s) installed during the program year are maintained in operating condition; and

(b) the extent to which there has been a significant reduction in the effectiveness of the measure(s).

MEASURED SAVINGS - Savings or reductions in billing determinants, which are determined using engineering analysis in combination with measured data or through billing analysis.

MEGAWATT (MW) - One thousand kilowatts (1,000 kW) or one million (1,000,000) watts. One megawatt is enough energy to power 1,000 average California homes.

MEGAWATT HOUR (MWh) - One thousand kilowatt-hours, or an amount of electricity that would supply the monthly power needs of 1,000 typical homes in the Western U.S. (This is a rounding up to 8,760 kWh/year per home based on an average of 8,549 kWh used per household per year. (U.S. DOE EIA, 1997 annual per capita electricity consumption figures.))

MESSAGE DILUTION FACTOR - The percent of a target market that is actually exposed to a message. A newspaper ad may reach two million people, but it may have a message dilution factor of .01 indicating that the message was actually seen by 20,000 people ($2,000,000 \times .01 = 20,000$).

METER - A device used to measure some quantity, which includes electrical demand, electrical energy, temperature, flow, etc. A device for measuring levels and volumes of a customer's gas or electricity use.

METERED DATA - Data collected at a customer's premises over time through a meter for a specific end use, or energy-using system (e.g., lighting and HVAC), or location (e.g., floors of a building or a whole premise). Metered data may be collected over a variety of time intervals. Usually refers to electricity or gas data.

METERED DEMAND - The average time rate of energy flow over a period of time recorded by a utility meter.

METERING - The collection of energy consumption data over time at a customer's premises through the use of meters. These meters may collect information about kWh, kW or therms, with respect to an end use, a circuit, a piece or equipment or a whole building (or facility). Short-term metering generally refers to data collection for no more than a few weeks. End use metering refers specifically to separate data collection for one or more end uses in a building, such as lighting, air conditioning or refrigeration. What is called "spot metering" is not metering in this sense, but is an instantaneous measurement (rather than over time) of volts, amps, watts or power factor to determine equipment size and/or power draw.

MODEL - A mathematical representation or calculation procedure that is used to predict the energy use and demand in a building or facility or to estimate efficiency program

savings estimates. Models may be based on equations that specifically represent the physical processes or may be the result of statistical analysis of energy use data.

MONITORING (equipment or system) - Gathering of relevant measurement data over time to evaluate equipment or system performance, e.g., chiller electric demand, inlet evaporator temperature and flow, outlet evaporator temperature, condenser inlet temperature, and ambient dry-bulb temperature and relative humidity or wet-bulb temperature, for use in developing a chiller performance map (e.g., kW/ton vs. cooling load and vs. condenser inlet temperature).

MULTICOLLINEARITY - When two or more independent variables in a regression model are highly correlated with each other producing high standard errors for the regression parameter. The mathematics of a regression model fail if there is perfect collinearity, an exact linear relationship between two or more independent variables. If the correlation between independent variables is higher than either has with the dependent variable, the problems of multicollinearity are highly likely.

NATURAL CHANGE - The change in base usage over time. Natural change represents the effects of energy-related decisions that would have been made in the absence of the utility programs by both program participants and non-participants.

NEES (or NEBS) - See NON-ENERGY EFFECTS.

NET LOAD IMPACT - The total change in load that is attributable to the utility DSM program. This change in load may include, implicitly or explicitly, the effects of free drivers, free riders, state or federal energy efficiency standards, changes in the level of energy service, and natural change effects.

NET-TO-GROSS RATIO - A factor representing net program load impacts divided by gross program load impacts that is applied to gross program load impacts to convert them into net program load impacts. This factor is also sometimes used to convert gross measure costs to net measure costs.

NEW CONSTRUCTION - Residential and non-residential buildings that have been newly built or have added major additions subject to Title 24, the California building standards code.

NON-DEPLETABLE ENERGY SOURCES - Energy that is not obtained from depletable energy sources. (See California Code of Regulations, Title 24, Section 2-5302.)

NON-ENERGY EFFECTS (NEES) or NON-ENERGY BENEFITS (NEBS) – The identifiable and sometimes quantifiable non-energy results associated with program implementation or participation. Some examples of NEEs include: reduced emissions & environmental benefits, productivity improvements, jobs created, reduced utility debt and

disconnects, and higher comfort and convenience level of participant. The effects of an energy efficiency or resource acquisition program that are other than energy saved.

NON-PARTICIPANT - Any customer who was eligible but did not participate in the utility program under consideration in a given program year.

NON-RESIDENTIAL - Facilities used for business, commercial, agricultural, institutional, and industrial purposes.

NON-RESIDENTIAL BUILDING - Any building which is heated or cooled in its interior, and is of an occupancy type other than Type H, I, or J, as defined in the Uniform Building Code, 1973 edition, as adopted by the International Conference of Building Officials.

NON-RESIDENTIAL HARD TO REACH - Those customers who do not have easy access to program information or generally do not participate in energy efficiency programs due to a language, business size, geographic, or lease (split incentive) barrier.

NORMALIZATION - Adjustment of the results of a model due to changes in baseline assumptions (non-independent variables) during the test or post-retrofit period.

NOx - Oxides of nitrogen that are a chief component of air pollution that can be produced by the burning of fossil fuels. Also called nitrogen oxides.

NUG - A non-utility generator. A generation facility owned and operated by an entity who is not defined as a utility in that jurisdictional area.

OBLIGATION TO SERVE - The obligation of a utility to provide electric service to any customer who seeks that service, and is willing to pay the rates set for that service. Traditionally, utilities have assumed the obligation to serve in return for an exclusive monopoly franchise.

OCCUPANCY SENSOR - A control device that senses the presence of a person in a given space, commonly used to control lighting systems in buildings.

OHM - A unit of measure of electrical resistance. One volt can produce a current of one ampere through a resistance of one ohm.

ORIENTATION - The position of a building relative to the points of a compass.

OVERARCHING EVALUATION STUDIES - Collection and analysis of information at a statewide or regional level that can be used for program and portfolio evaluation planning and policy decision-making purposes.

PARALLEL SURVEY DESIGN - A survey in which similar questions on the same topic are asked to several distinct groups to assess construct validity.

PARTICIPANT - An individual, household, business, or other utility customer that received the service or financial assistance offered through a particular utility DSM program, set of utility programs, or particular aspect of a utility program in a given program year. Participation is determined in the same way as reported by a utility in its Annual DSM Summary.

PARTICIPANT TEST - A cost-effectiveness test intended to measure the cost-effectiveness of energy efficiency programs from the perspective of electric and/or gas customers (individuals or organizations) participating in them.

PARTIES OR INTERESTED PARTIES - Persons and organizations with an interest in energy efficiency that comment on or participate in the Commission's efforts to develop and implement ratepayer-funded energy efficiency programs.

PASSIVE SOLAR ENERGY - Use of the sun to help meet a building's energy needs by means of architectural design (such as arrangement of windows) and materials (such as floors that store heat).

PASSIVE SOLAR SYSTEM - A solar heating or cooling system that uses no external mechanical power to move the collected solar heat.

PBR - See PERFORMANCE-BASED REGULATION.

PEAK DEMAND - The maximum level of metered demand during a specified period, such as a billing month, or during a specified peak demand period.

PEAK DEMAND PERIOD - Noon to 7 p.m. Monday through Friday, June, July, August, and September.

PEAK LOAD - The highest electrical demand within a particular period of time. Daily electric peaks on weekdays occur in late afternoon and early evening. Annual peaks occur on hot summer days.

PERFORMANCE CONTRACTS - A binding agreement between two parties prescribing the range and magnitude of achievement required of equipment, subsystem, or system, which is provided by one party for the benefit and use of the other.

PERFORMANCE MANAGEMENT - The determination or the extent to which a person, organization, or program is successfully meeting specified goals and objectives.

PERFORMANCE-BASED REGULATION (PBR) - Any rate-setting mechanism which attempts to link rewards (generally profits) to desired results or targets. PBR sets rates, or components of rates, for a period of time based on external indices rather than a utility's cost-of-service. Other definitions include light-handed regulation which is less costly and

less subject to debate and litigation. A form of rate regulation which provides utilities with better incentives to reduce their costs than cost-of-service regulation.

PERSISTENCE STUDY - A study to assess changes in net program impacts over time (that include retention and technical degradation).

PGC - See PUBLIC GOODS CHARGE.

PORTFOLIO - All IOU and non-IOU energy efficiency programs funded through the PGC that are implemented during a program year or cycle.

POST RETROFIT PERIOD - The time following a retrofit during which savings are to be determined.

PRACTICE - Generally refers to a change in a customer's behavior or procedures that reduces energy use (e.g., thermostat settings, maintenance procedures).

PRACTICE RETENTION STUDY - An assessment of the length of time a customer continues the energy conservation behavioral changes after adoption of these changes.

PRECISION - The indication of the closeness of agreement among repeated measurements of the same physical quantity. In econometrics, the accuracy of an estimator as measured by the inverse of its variance.

PROCESS EVALUATION - A systematic assessment of an energy efficiency program for the purposes of (a) documenting program operations at the time of the examination, and (b) to identify and recommend improvements that can be made to the program to increase the program's efficiency or effectiveness for acquiring energy resources while maintaining high levels of participant satisfaction.

PROCESS OVERHAUL - Modifications to industrial or agricultural processes to improve their energy use characteristics.

PROGRAM - An activity, strategy, or course of action undertaken by an implementer or administrator using PGC funds. Each program is defined by a unique combination of program strategy, market segment, marketing approach, and energy efficiency measure(s) included.

PROGRAM (IMPLEMENTATION) CYCLE - The period of time over which programs are funded, planned and implemented. Can be an annual cycle, a bi-annual cycle or other period of time.

PROGRAM DESIGN - The method or approach for making, doing, or accomplishing an objective by means of a program.

PROGRAM DEVELOPMENT - The process by which ideas for new or revised energy efficiency programs are converted into a design to achieve a specific objective.

PROGRAM PENETRATION - The level of program acceptance among qualified customers.

PROGRAM MANAGEMENT - The responsibility and ability to oversee and guide the performance of a program to achieve its objective.

PROGRAM STRATEGIES - Refers to the type of method deployed by the program in order to obtain program participation. Some examples of program strategies include: rebates, codes, performance contracting and audits.

PROGRAM THEORY - A presentation of the goals of a program, incorporated with a detailed presentation of the activities that the program will use to accomplish those goals and the identification of the causal relationships between the activities and the program's effects.

PROGRAM YEAR - The calendar year in which program participation occurs.

PROGRAMMABLE CONTROLLER - A device that controls the operation of electrical equipment (such as air conditioning units and lights) according to a preset time schedule.

PROJECT - An activity or course of action undertaken by an implementer involving one or multiple energy efficiency measures, usually at a single site.

PROJECT DEVELOPMENT - The process by which an implementer identifies a strategy or creates a design to provide energy efficiency products, services, and practices directly to customers.

PUBLIC GOODS CHARGE (PGC) - Per Assembly Bill (AB) 1890, a universal charge applied to each electric utility Customer's bill to support the provision of public goods. Public goods covered by California's electric PGC include public purpose energy efficiency programs, low-income services, renewables, and energy-related research and development.

PUBLIC INTEREST GOALS - Public interest goals of electric utility regulation include: (a) inter-and intra-class and intergenerational equity; (b) the equal treatment of equals (horizontal equity); (c) balancing long- and short-term goals that have the potential to affect intergenerational balance; (d) protecting against the abuse of monopoly power; and (e) general protection of the health and welfare of the citizens of the state, nation, and world. Environmental and other types of social costs are subsumed under the equity and health and welfare responsibilities.

RADIANT BARRIER - A device designed to reduce or stop the flow of radiant energy.

RATIO ESTIMATOR (SAMPLING METHOD) - A sampling method to obtain increased precision by taking advantage of the correlation between an auxiliary variable and the variable of interest to reduce the coefficient of variation.

REBATES - A type of incentive provided to encourage the adoption of energy efficient practices, typically paid after the measure has been installed. There are typically two types of rebates: a Prescriptive Rebate, which is a prescribed financial incentive per unit for a prescribed list of products, and a Customized Rebate, in which the financial incentive is determined using an analysis of the customer's equipment and an agreement on the specific products to be installed. Upstream rebates are financial incentives provided for manufacturing, sales, stocking, or other per unit energy efficient product movement activities designed to increase use of particular type of products.

REBOUND EFFECT - A change in energy using behavior that yields an increased level of service and that occurs as a result of taking an energy efficiency action.

RECALL - The ability to remember an event and discuss reactions to the event. Typically used in market effects evaluations in which a number of people may have been exposed to a promotional message, but only a small percent may remember (or recall) the message and be able to discuss the influence of the message.

RECALL RESPONSE RATE - The percent of a target market exposed to a message that recalls the message.

REGRESSION MODEL - A mathematical model based on statistical analysis where the dependent variable is regressed on the independent variables which are said to determine its value. In so doing, the relationship between the variables is estimated statistically from the data used.

RELIABILITY - When used in energy evaluation, refers to the likelihood that the observations can be replicated.

REMODELING - Modifications to the characteristics of an existing residential or non-residential building or energy-using equipment installed within it.

RENEWABLE ENERGY - Resources that constantly renew themselves or that are regarded as practically inexhaustible. These include solar, wind, geothermal, hydro and wood. Although particular geothermal formations can be depleted, the natural heat in the earth is a virtually inexhaustible reserve of potential energy. Renewable resources also include some experimental or less-developed sources such as tidal power, sea currents and ocean thermal gradients.

RENEWABLE RESOURCES - Renewable energy resources are naturally replenishable, but flow-limited. They are virtually inexhaustible in duration but limited in the amount of energy that is available per unit of time. Some (such as geothermal and biomass) may be stock-limited in that stocks are depleted by use, but on a time scale of decades, or perhaps

centuries, they can probably be replenished. Renewable energy resources include: biomass, hydro, geothermal, solar and wind. In the future they could also include the use of ocean thermal, wave, and tidal action technologies. Utility renewable resource applications include bulk electricity generation, on-site electricity generation, distributed electricity generation, non-grid-connected generation, and demand-reduction (energy efficiency) technologies.

RENOVATION - Modifications to the characteristics of an existing residential or non-residential building, including but not limited to windows, insulation, and other modifications to the building shell.

REPLACEMENT - Refers to the changing of equipment either due to failure, move to more efficient equipment or other reasons near the end of product life or earlier. Often used to refer to a move to a more energy efficient product that replaces an inefficient product.

RESEARCH AND DEVELOPMENT (R&D) - Research is the discovery of fundamental, new knowledge. Development is the application of new knowledge to develop a potential new service or product. Basic power sector R&D is most commonly funded and conducted through the Department of Energy (DOE), its associated government laboratories, university laboratories, the Electric Power Research Institute (EPRI), and private sector companies.

RESIDENTIAL BUILDING - Means any hotel, motel, apartment house, lodging house, single dwelling, or other residential building that is heated or mechanically cooled.

RESIDENTIAL CUSTOMER - Existing single family residences, multi-family dwellings (whether master-metered or individually metered), and buildings that are essentially residential but used for commercial purposes, including but not limited to time shares and vacation homes.

RESIDENTIAL HARD-TO-REACH - Those customers who do not have easy access to program information or generally do not participate in energy efficiency programs due to a language, income, housing type, geographic, or home ownership (split incentives) barrier.

RETAIL MARKET - A market in which electricity and other energy services are sold directly to the end use customer.

RETENTION (MEASURE) - The degree to which measures are retained in use after they are installed.

RETROFIT - Energy efficiency activities undertaken in existing residential or non-residential buildings where existing inefficient equipment is replaced by efficient equipment.

RETROFIT ISOLATION - The savings measurement approach defined in the International Performance Measurement and Verification Protocols (IPMVP) Options A and B, and ASHRAE Guideline 14 that determines energy or demand savings through the use of meters to isolate the energy flows for the system(s) under consideration.

ROADMAP - Set of decision trees or decision flow diagrams that support the process of determining if an evaluation of the program is necessary, and what type of evaluations, methods or steps can be used.

R-VALUE - A unit of thermal resistance used for comparing insulating values of different material. It is basically a measure of the effectiveness of insulation in stopping heat flow. The higher the R-value number of a material, the greater its insulating properties and the slower the heat flow through it. The specific value needed to insulate a home depends on climate, type of heating system, and other factors.

SAE - See STATISTICALLY ADJUSTED ENGINEERING MODELS.

SAMPLE DESIGN - The approach used to select the sample units.

SAMPLING ERROR - The error in estimating a parameter caused by the fact that in the sample at hand all the disturbances are not zero.

SAVINGS DETERMINATION - The process of separating a retrofit's (energy efficiency measure's) effectiveness from a facility's energy use pattern. It involves measurements of physical conditions and analysis of resultant data.

SAVINGS MEASUREMENT APPROACH - The estimation of energy and demand savings associated with an energy efficiency measure for a piece of equipment, a subsystem, or a system. The estimated savings are based on some kind of measured data from before and after the retrofit and may be calculated using a variety of engineering techniques.

SEASONAL ENERGY EFFICIENCY RATIO (SEER) - The total cooling output of a central air conditioning unit in Btus during its normal usage period for cooling divided by the total electrical energy input in watt-hours during the same period, as determined using specified federal test procedures. (See Title 20, Section 2-1602(c)(11).)

SERIAL CORRELATION - See AUTOCORRELATION.

SENSIBLE HEAT - Heat that results in a temperature change.

SERVICE AREA - The geographical territory served by a utility.

SETBACK THERMOSTAT - See THERMOSTAT, SETBACK.

SHADE SCREEN - A screen affixed to the exterior of a window or other glazed opening designed to reduce the solar radiation reaching the glazing.

SHADING - (a) The protection from heat gains due to direct solar radiation; (b) Shading is provided by (1) permanently attached exterior devices, glazing materials, adherent materials applied to the glazing, or an adjacent building for non-residential buildings, hotels, motels and high rise apartments, and by (2) devices affixed to the structure for residential buildings. (See California Code of Regulations, Title 24, Section 2-5302.)

SHADING COEFFICIENT (SC) - The ratio of solar heat gain through fenestration, with or without integral shading devices, to that occurring through unshaded 1/8 in. thick clear double strength glass. See also **SOLAR HEAT GAIN COEFFICIENT**.

SHGC - See **SOLAR HEAT GAIN COEFFICIENT**.

SIMPLE RANDOM SAMPLING - A method of selecting n sample units out of the N population such that every one of the distinct N items has an equal chance of being selected.

SIMPLIFIED ENGINEERING MODEL - Engineering equations used to calculate energy usage and/or savings. These models are usually based on a quantitative description of physical processes that describe the transformation of delivered energy into useful work such as heat, lighting or motor drive. In practice, these models may be reduced to simple equations that calculate energy usage or savings as a function of measurable attributes of customers, facilities or equipment (e.g., lighting use = watts X hours of use). These models do not incorporate billing data and do not produce estimates of energy savings to which tests of statistical validity can be applied.

SOLAR HEAT GAIN - Heat added to a space due to transmitted and absorbed solar energy.

SOLAR HEAT GAIN COEFFICIENT (SHGC) - The ratio of the solar heat gain entering the space through the fenestration area to the incident solar radiation.

SOLAR HEAT GAIN FACTOR - An estimate used in calculating cooling loads of the heat gain due to transmitted and absorbed solar energy through 1/8"-thick, clear glass at a specific latitude, time, and orientation.

SOLAR HEATING AND HOT WATER SYSTEMS - Solar heating or hot water systems provide two basic functions: (a) capturing the sun's radiant energy, converting it into heat energy, and storing this heat in insulated storage tank(s); and (b) delivering the stored energy as needed to either the domestic hot water or heating system. These components are called the collection and delivery subsystems.

SPECIAL CONTRACTS - Any contract that provides a utility service under terms and conditions other than those listed in the utility's tariffs. For example, an electric utility

may enter into an agreement with a large customer to provide electricity at a rate below the tariffed rate in order to prevent the customer from taking advantage of some other option that would result in the loss of the customer's load. This generally allows that customer to compete more effectively in their product market.

SPILOVER - Reductions in energy consumption and/or demand in a utility's service area caused by the presence of the DSM program, beyond program related gross savings of participants. These effects could result from: (a) additional energy efficiency actions that program participants take outside the program as a result of having participated; (b) changes in the array of energy-using equipment that manufacturers, dealers, and contractors offer all customers as a result of program availability; and (c) changes in the energy use of non-participants as a result of utility programs, whether direct (e.g., utility program advertising) or indirect (e.g., stocking practices such as (b) above, or changes in consumer buying habits).

SPLIT-THE-SAVINGS (Electric Utility) - The basis for settling economy-energy transactions between utilities. The added costs of the supplier are subtracted from the avoided costs of the buyer, and the difference is evenly divided.

SPURIOUSNESS OR SPURIOUS CORRELATION - The apparent association between two variables that is actually attributable to a third variable outside the current analysis, probably a common precedent variable.

STAKEHOLDERS - In program evaluation, stakeholders refer to the myriad of parties that are impacted by a program. Stakeholders include: regulatory staff, program designers, implementers and evaluators, energy producers, special interest groups, potential participants and customers.

STANDARD DEVIATION - The square root of the variance.

STANDARD PERFORMANCE CONTRACT (SPC) - Programs consisting of a set of agreements between the administrator or implementer and a number of project sponsors (either implementers or customers) to deliver energy savings from the installation of energy efficiency measures and technologies at a facility or set of facilities. These agreements are for a pre-specified price per unit of energy savings, measured using a pre-specified set of measurement and verification (M&V) protocols. An SPC program is an open-ended offer with a pre-specified price and set of terms.

STANDBY LOSS - A measure of the losses from a water heater tank. When expressed as a percentage, standby loss is the ratio of heat loss per hour to the heat content of the stored water above room temperature. When expressed in watts, standby loss is the heat lost per hour, per square foot of tank surface area. (See California Code of Regulations, Title 20, Section 1602(f)(5).)

STATEWIDE MARKETING AND OUTREACH PROGRAMS - Programs that convey consistent statewide messages to individual consumers through a mass-market advertising campaign.

STATEWIDE PROGRAM - A program available in the service territories of all four large IOUs, with identical implementation characteristics in all areas, including incentives and application procedures.

STATISTICAL ANALYSIS - Extrapolation of sample data up to the population, calculation of error bounds.

STATISTICAL COMPARISONS - A comparison group of customers serving as a proxy of what program participants would have looked like if the program had not been offered.

STATISTICALLY ADJUSTED ENGINEERING (SAE) MODELS - A category of billing analysis models that incorporate the engineering estimate of savings as a dependent variable. The regression coefficient in these models is the percentage of the engineering estimate of savings observed in changes in energy usage. For example, if the coefficient on the SAE term is 0.8, this means that the customers are on average realizing 80% of the savings from their engineering estimates.

STEADY-STATE EFFICIENCY - A performance rating for space heaters; a measure of the percentage of heat from combustion of gas which is transferred to the space being heated under specified steady-state conditions. (See California Code of Regulations, Title 20, Section 1602(e)(13).)

STRANDED BENEFITS - Public interest programs and goals which could be compromised or abandoned by a restructured electric industry. These potential “stranded benefits” might include: environmental protection, fuel diversity, energy efficiency, low-income ratepayer assistance, and other types of socially beneficial programs.

STRATIFIED RANDOM SAMPLING – The population is divided into X units of subpopulations that are non-overlapping and together comprise the entire population, called strata. A simple random sample is taken of each strata to create a sample based upon stratified random sampling.

STRATIFIED RATIO ESTIMATION - A sampling method that combines a stratified sample design with a ratio estimator to reduce the coefficient of variation by using the correlation of a known measure for the unit (e.g., expected energy savings) to stratify the population and allocate sample from strata for optimal sampling.

SUNK COST - In economics, a sunk cost is a cost that has already been incurred, and therefore cannot be avoided by any strategy going forward.

SUPPLY-SIDE - Activities conducted on the utility’s side of the customer meter. Activities designed to supply electric power to customers, rather than meeting load

though energy efficiency measures or on-site generation on the customer side of the meter.

SYSTEM - A combination of equipment and/or controls, accessories, interconnecting means, and terminal elements by which energy is transformed so as to perform a specific function, such as HVAC, service water heating, or illumination.

SYSTEM INTEGRATION (OF NEW TECHNOLOGIES) - The successful integration of a new technology into the electric utility system by analyzing the technology's system effects and resolving any negative impacts that might result from its broader use.

TECHNICAL DEGRADATION FACTOR - A multiplier used to account for time- and use-related change in the energy savings of a high efficiency measure or practice relative to a standard efficiency measure or practice.

TECHNICAL POTENTIAL - The complete penetration of all measures analyzed in applications where they were deemed technically feasible from an engineering perspective.

TEMPERATURE - Degree of hotness or coldness measured on one of several arbitrary scales based on some observable phenomenon (such as the expansion).

THERM - One hundred thousand (100,000) British thermal units (1 therm = 100,000 Btu).

THERMAL BREAK (thermal barrier) - An element of low heat conductivity placed in such a way as to reduce or prevent the flow of heat. Some metal framed windows are designed with thermal breaks to improve their overall thermal performance.

THERMAL CONDUCTANCE (C) - The constant time rate of heat flow through unit area of a body induced by a unit temperature difference between the surfaces, Btu/(ft²-h-°F) or W/(m²-K). It is the reciprocal of thermal resistance. See THERMAL RESISTANCE.

THERMAL MASS - A material used to store heat, thereby slowing the temperature variation within a space. Typical thermal mass materials include concrete, brick, masonry, tile and mortar, water, and rock or other materials with high heat capacity.

THERMAL RESISTANCE (R) - The reciprocal of thermal conductance; 1/C as well as 1/h, 1/U, h-ft²-°F/Btu.

THERMAL (ENERGY) STORAGE - A technology that lowers the amount of electricity needed for comfort conditioning during utility peak load periods. A buildings thermal energy storage system might, for example, use off-peak power to make ice or to chill water at night, later using the ice or chilled water in a power saving process for cooling during the day. See THERMAL MASS.

THERMOSTAT - An automatic control device designed to be responsive to temperature and typically used to maintain set temperatures by cycling the HVAC system.

THERMOSTAT, SETBACK - A device containing a clock mechanism, which can automatically change the inside temperature maintained by the HVAC system according to a preset schedule. The heating or cooling requirements can be reduced when a building is unoccupied or when occupants are asleep. (See California Code of Regulations, Title 24, Section 2- 5352(h).)

TIME-OF-USE METER - A measuring device that records the times during which a customer uses various amounts of electricity. This type of meter is used for customers who pay time-of-use rates.

TIME-OF-USE (TOU) RATES - Electricity prices that vary depending on the time periods in which the energy is consumed. In a time-of- use rate structure, higher prices are charged during utility peak-load times. Such rates can provide an incentive for consumers to curb power use during peak times.

TOTAL RESOURCE COST TEST – SOCIETAL VERSION - A cost-effectiveness test intended to measure the overall cost-effectiveness of energy efficiency programs from a societal perspective.

TOU - See TIME OF USE RATES.

TRIANGULATION - Comparing the results from two or more different data gathering or measurement techniques on the same problem to derive a “best” estimate from the analysis of the comparison.

UA - A measure of the amount of heat that would be transferred through a given surface or enclosure (such as a building envelope) with a one degree Fahrenheit temperature difference between the two sides. The UA is calculated by multiplying the U-value by the area of the surface (or surfaces).

UNCERTAINTY - The range or interval of doubt surrounding a measured or calculated value within which the true value is expected to fall within some degree of confidence.

UNCERTAINTY ANALYSIS - (a) A procedure or method by which the uncertainty of a measured or calculated value is determined; (b) the process of determining the degree of confidence in the true value when using a measurement procedure(s) and/or calculation(s).

UNCONDITIONED SPACE - A space that is neither directly nor indirectly conditioned space, which can be isolated from conditioned space by partitions and/or closeable doors. (See California Code of Regulations, Title 24, Section 2-5302.)

UNIVERSAL SERVICE - Electric service sufficient for basic needs (an evolving bundle of basic services) available to virtually all members of the population regardless of income.

UPGRADE (Electric utility) - Replacement or addition of electrical equipment resulting in increased generation or transmission capability.

UPSTREAM PROGRAMS - Programs that provide information and/or financial assistance to entities in the delivery chain of high-efficiency products at the retail, wholesale, or manufacturing level.

UTILITY METER - The meter used to calculate a monthly energy and/or demand charge at a specific utility/customer connection; more than one may be installed per customer and per site due to different supply voltages, capacity requirements, physical separation distances, installation periods, or for specific customer requirements or utility programs.

U-VALUE/U-FACTOR - A measure of how well heat is transferred by the entire window - the frame, sash and glass - either into or out of the building. U-value is the opposite of R-value. The lower the U-factor number, the better the window will keep heat inside a home on a cold day.

VARIABLE AIR VOLUME (VAV) HVAC system - HVAC systems that control the dry-bulb temperature within a space by varying the volume of supply air to the space.

VENTILATION - The process of supplying or removing air by natural or mechanical means to or from any space. Such air may or may not have been conditioned or treated.

VERIFICATION PHASE - The step of the California Energy Contingency Plan to determine the existence and scope of an energy shortage and report to Energy Commission executives, the Governor and the Legislature where required under the plan.

WATT - A unit of measure of electric power at a point in time, as capacity or demand. One watt of power maintained over time is equal to one joule per second. Some Christmas tree lights use one watt. The watt is named after Scottish inventor James Watt and is capitalized when shortened to W and used with other abbreviations, as in kWh.

WATT-HOUR - One watt of power expended for one hour. One thousandth of a kilowatt-hour.

WEATHERSTRIPPING - Specially designed strips, seals and gaskets installed around doors and windows to limit air leakage.

WET-BULB TEMPERATURE - The temperature at which water, by evaporating into air, can bring the air to saturation at the same temperature. Wet-bulb temperature is measured by a wet-bulb psychrometer.

WHOLE-BUILDING CALIBRATED SIMULATION APPROACH - The savings measurement approach defined in the International Performance Measurement and Verification Protocols (IPMVP) Option D and ASHRAE Guideline 14, which involves the use of an approved computer simulation program to develop a physical model of the building in order to determine energy and demand savings. The simulation program is used to model the energy used by the facility before and after the retrofit. The pre- or post-retrofit models are developed by calibration with measured energy use and demand data and weather data.

WHOLE-BUILDING METERED APPROACH - The savings measurement approach defined in the International Performance Measurement and Verification Protocols (IPMVP) Option C and ASHRAE Guideline 14 that determines energy and demand savings through the use of whole-facility energy (end use) data, which may be measured by utility meters or data loggers. This approach may involve the use of monthly utility billing data or data gathered more frequently from a main meter.

ZONE - A space or group of spaces within a building with any combination of heating, cooling, or lighting requirements sufficiently similar so that desired conditions can be maintained throughout by a single controlling device.

APPENDIX B: Bibliography

- ADM Associates and TecMRKT Works (2000). *Statewide Survey of Multi-Family Common Area Building Owners Market: Volume 1: Apartment Complexes*. Study ID 3514.
- Aigner, Dennis and Joseph Hirschberg (1985). "Commercial/industrial customer response to time-of-use electricity prices: some experimental results." *Rand Journal of Economics* 16 (3): 341-355.
- Aigner, Dennis and Lee Lillard (1984). "Measuring Peak Load Pricing Response from Experimental Data: An Exploratory Analysis." *Journal of Business & Economic Statistics* 2 (1): 21-39.
- AIS, SRC International (2001). *European Ex-Post Evaluation Guidebook for DSM and EE Services Programmes*. International Energy Agency. April.
- Amalfi, John, Pete Jacobs and Roger Wright (1996). "Short-Term Monitoring of Commercial Lighting Systems - Extrapolation from the Measurement Period to Annual Consumption." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA.
- Amemiya, Takeshi and Thomas McCurdy (1986). "Instrumental-Variable Estimation of an Error-Components Model." *Econometrica* 54 (4): 869-880.
- American Statistical Association (1995). What is a Survey?
<<http://www.amstat.org/sections/srms/>>
- Anstett, Michael and Jan Kreider (1992). "Application of Neural Networking Models to Predict Energy Use." *ASHRAE Transactions* 98 (2): 505-516.
- ASHRAE (1999). Chapter 39: Building Energy Monitoring. *ASHRAE Handbook of Fundamentals*. ASHRAE: Atlanta, GA.
- ASHRAE (2001a). *ASHRAE Standard 14-2001 - Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs*. ASHRAE.
- ASHRAE (2001b). Cooling Load Calculations (Chapters 28 and 29). *ASHRAE Handbook of Fundamentals*. ASHRAE: Atlanta, GA. I-P Edition.
- ASHRAE (2001c). *Method of Test for Determining Design and Seasonal Efficiency of Residential Thermal Distribution Systems, Third Public Review Draft*. American Society of Heating, Refrigeration and Air Conditioning Engineers. ASHRAE Standard 152-P.
- * ASHRAE (2002). Measurement of Energy and Demand Savings, Guideline 14. American Society of Heating, Refrigeration and Air Conditioning Engineers: Atlanta, GA.
- ASW Engineering Management Consultants Inc. and Megdal & Associates (2001). *Measure Retention Study of the 1996 Commercial Energy Efficiency Incentive (EEI) Program*. Southern California Gas Company. February. Study ID 720.
- * Auch, Lynn and Mike McDonald (1994). "Conservation Advertising Campaigns and Advertising Effectiveness Research: The Right Combination to Solidify the Conservation Ethic." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 1.1-1.7.

* Indicates that the information was used in the construction of the Framework.

- Balestra, Pietro and Marc Nerlove (1966). "Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas." *Econometrica* 34 (3): 585-612.
- Blake, William, Shel Feldman and Dorothy Conant (2003). "Dressing the Priestess: Preparation for and Results of a Delphi Study for a Residential New Construction Program." *International Energy Program Evaluation Conference*: Seattle, WA. pp. 863-874.
- Blalock, Herbert M (1979). *Social Statistics*. 2nd Edition. McGraw-Hill: New York, NY.
- Blumstein, Carl, Seymour Goldstone and Loren Lutzenhiser (2000). "A theory-based approach to market transformation." *Energy Policy* 28: 137-144.
- * Bordner, Robert , Mark Siegal and Lisa Skumatz (1994). "The Application of Survival Analysis to Demand-Side Management Evaluation." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 8.9-8.18.
- Bronfman, Benson and Jane Peters (1991). *Process Evaluations of DSM Programs*. Oak Ridge National Laboratory. December.
- * Brost, Matt , Roger Wright, Ramona Peet, Cathy Chappell, Douglas Mahone and Pete Jacobs (2002). "Measuring Accomplishments of Energy Efficiency in California's Non-Residential New Construction Market." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 10.35-10.45.
- Buller, Susan, William Miller, Richard Barnes, Fred Coito and Ilene Obstfeld (1993). "Combining Monitoring, Engineering Analysis, and Billing Analysis to Evaluate PG&E's Commercial Retrofit Incentive Program." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 265-272.
- * Buller, Susan, Glen Weisbord, Kenneth Train and Richard Barnes (1994). "Modeling Technology Adoption and DSM Program Decisions." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 1.9-1.17.
- * CADMAC (1998). *Quality Assurance Guidelines for Statistical, Engineering, and Self-Report Methods for Estimating DSM Program Impacts*. Appendix J.
- * CADMAC (1999). *Protocols and Procedures for the Verification of Costs, Benefits, and Shareholder Earnings from Demand-Side Management Programs*. California Public Utilities Commission. 93-05-063.
- * California State Governor's Office (2001). *Standard Practice Manual (SPM): Economic Analysis of Demand-Side Management Programs*. October.
- Cambridge Systematics Inc. and Freeman Sullivan and Company (1994). *DSM Free Ridership Study*. Empire State Electric Energy Research Corporation. November. Report EP-92-65.
- Campbell, Donald and Julian Stanley (1963). *Experimental and Quasi-Experimental Designs for Research*. Houghton Mifflin: Boston, MA.
- Canadian Health Services Research Foundation (2000). *Health Services Research and Evidence-Based Decision-Making*. June.
- Cavalli, John, Christie Torok and Valerie Richardson (1999). "Self-Reports and Market Transformation: A Compelling New Approach." *International Energy Program Evaluation Conference*: Denver, CO. pp. 773-787.

* Indicates that the information was used in the construction of the Framework.

- CBEE Technical Services Consultants (1998). *Proposed Guidelines for Conducting Market Assessment and Evaluation*.
- * Cochran, William G (1977). *Sampling Techniques*. 3rd Edition. John Wiley & Sons: New York, NY.
- Cohen, Jacob (1990). *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, Inc.: Englewood Cliffs, NJ.
- * Coito, Fred and Richard Barnes (1996). "Improving Billing Analysis Results Using On-Site Follow-up Surveys." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 6.35-6.39.
- Coito, Fred and Mike Rufo (2002). *California Statewide Commercial Sector Energy Efficiency Potential Study*. Pacific Gas & Electric. May 14. Study ID SW061.
- Conant, Dorothy and Jeremy Schutte (1993). "Matching Methodologies and Program Types: An Evaluation Retrospective." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 155-160.
- * Conlon, Tom, Glen Weisbord and Shahanna Samiullah (1999). "How Can We Tell if Free Information is Really Transforming Our Market?" *International Energy Program Evaluation Conference*: Denver, CO. pp. 829-840.
- Cook, Thomas (2000). "The False Choice Between Theory-Based Evaluation and Experimentation." *New Directions for Evaluation* Fall (87): 27-34.
- Cornwell, Christopher and Peter Rupert (1988). "Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variables Estimators." *Journal of Applied Econometrics* 3: 149-155.
- CPUC (2001). *Energy Efficiency Policy Manual*. California Public Utilities Commission. October.
- CPUC (2003). Assigned Commissioner's Ruling Proposing Direction and Scope for Further Rulemaking. California Public Utilities Commission: San Francisco, CA.
- * CPUC (2003). *Energy Efficiency Policy Manual, Version 2*. California Public Utility Commission. August.
- Davidson, Jane E. (2000). "Ascertaining Causality in Theory-Based Evaluation." *New Directions for Evaluation* (Number 87): 17-26.
- DeCotis, Paul A, Mark C Coleman, Jennifer Ellefsen and Helen Kim (2000). "Portfolio Approach to Designing and Evaluating Buildings Energy Efficiency Programs." *National Energy Services Conference*: Lake Worth, FL. pp. 495-503.
- DeCotis, Paul A. and John F. Munro (2001). "A Theory-Based Systems Approach for Evaluating Energy Efficiency Investments Portfolios." *National Energy Services Conference*: Lake Worth, FL. pp. 282-296.
- DeCotis, Paul A, Larry Pakenas, and Joel Eisenberg (2002). "Systems-Based Portfolio Evaluation: Diagnostic Benefits and Methodological Challenges." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 10.57-10.67.
- * Dohrmann, Donald R., Marian V. Brown and Martin H. Morse (1999). "A Longitudinal Study of Non-Residential DSM Measure Retention." *International Energy Program Evaluation Conference*: Denver, CO. pp. 695-705.

* Indicates that the information was used in the construction of the Framework.

- England, Paula, George Farkas and Margaret Barton (1988). "Explaining Occupational Sex Segregation and Wages: Findings From a Model with Fixed Effects." *American Sociological Review* 53: 544-558.
- Energy and Environmental Economics (2004). *A Forecast of Cost Effectiveness Avoided Costs and Externality Adders*. San Francisco, CA.
- * Engle, Victoria, Lori Megdal, Tom Rooney, Lawrence Pakenas and Susan Sowek (2003). "Quantifying Load-Shifting Benefits from an Advertising Campaign." *International Energy Program Evaluation Conference*: Seattle, WA. pp. 393-401.
- * EPRI (1992). *DSM Process Evaluation: A Guidebook To Current Practice*. Electric Power Research Institute. Project RP2981-2.
- Erickson, Jeff and Oscar Bloch (2002). "Using Program Theories to Align Performance Metrics with Public Purpose Goals." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 5.65-5.75.
- * Eto, Joe, Ralph Pahl and Jeff Schlegel (1996). *A Scoping Study on Energy-Efficiency Market Transformation by California Utility DSM Programs*. Ernest Orlando Lawrence Berkeley National Laboratory. July. LBNL-39059 UC-1322.
- Fels, Margaret, Miriam Goldberg and Michael Lavine (1986). "Exploratory Scorekeeping for Oil-Heated Houses." *Energy and Buildings* Vol 9 (1,2): 127-136.
- * FEMP (2000). *Federal Energy Management Program (FEMP) M&V Guidelines: Measurement and Verification for Federal Energy Projects*. Federal Energy Management Program. September. Version 2.2, DOE/GO-102000-0960.
- Field Diagnostic Services (2003). HVAC Service Assistant. <<http://www.acrx.com/>>
- Francisco, Paul W. and Larry S. Palmiter (2003). "Field Evaluation of a New Device to Measure Air Handler Flow." *ASHRAE Transactions* Vol 109, Part 2.
- Gabel-Dodd/Energy Soft LLC (2001). EnergyPro. Version 3.141. <www.energysoft.com>
- Global Energy Partners (2003). *California Summary Study of 2001 Energy Efficiency Programs*. March 13. Study ID 10019.
- Goldberg, Miriam (1986). "A Midwest Low-Income Weatherization Program Seen Through PRISM." *Energy and Buildings* Vol 9 (1,2): 37-44.
- Goldberg, Miriam (2003). "Does Talking About Barriers Just Get in the Way?" *International Energy Program Evaluation Conference*: Seattle, WA. pp. 145-155.
- Goldberg, Miriam and G Kennedy Agnew (2003). "Dynamic Modeling of Market Effects & Spillover with Limited Information." *National Energy Service Conference*: New Orleans, LA. pp. 4.1-4.15.
- Goldberg, Miriam, G Kennedy Agnew and Lori Boekeler (2003). *Business Programs Evaluation: Market Effects Pro Forma Estimates*. Wisconsin Department of Administration, Division of Energy.
- Goldberg, Miriam and Edmund Kademan (1995). "Is it Net or Not? A Simulation Study of Two Methods." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 459-465.
- Goldberg, Miriam and Kurt Scheuermann (1997). "Gross and Net Savings Analysis for Unique Projects." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 389-395.

* Indicates that the information was used in the construction of the Framework.

- Goldstone, Seymour, Mike Rufo and John Wilson (2000). "Applying a Theory-Based Approach to California's Non-Residential Standard Performance Contract Program: Lessons Learned." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 5.103-5.125.
- Green Building XML (gbXML) (2003). Geopraxis. 0.33. www.geopraxis.com
- * Green, Jerry and Lisa Skumatz (2000). "Evaluating the Impacts of Education/Outreach Programs: Lessons on Impacts, Methods, and Optimal Education." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 8.123-8.136.
- * Grover, Stephan, David Cohen and My K Ton (2002). "Saturation, Penetration, Transformation: How Do You Know When a Market Has Changed?" *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 10.99-10.110.
- Groves, Robert M, Floyd J Fowler, Mick P Couper, James Lepkowski, Eleanor Singer and Roger Tourangeau (2004). *Survey Methodology*. Wiley: Hoboken, NJ.
- Haberl, Jeff S, Atch Sreshthaputra, David E Claridge and John Kelley Kissock (2003). "Inverse Modeling Toolkit: Application and Testing." *ASHRAE Transactions Vol 109* (Part 2).
- Hagler Bailly Consulting (1998). *CTAC Market Effects Study*. Southern California Edison. Study 3504.
- Hall, Nick and Gretchen Jordan (2001). *The 2001 FEMP Customer Survey: Study Report*. Sandia National Laboratories.
- Hall, Nick and John Reed (1997a). "Methods for Measuring Customer Satisfaction." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 23-33.
- Hall, Nick and John Reed (1997b). *Residential Low-Income Energy Management Program: Process Evaluation Report*. Cinergy. December.
- Hall, Nick and John Reed (1998). *Process and Impact Evaluation of Missouri Gas Energy's Pilot Weatherization Program*. Missouri Gas Energy. March.
- Hall, Nick and John Reed (1999). "Market Transformation: Expectations vs. Reality." *Home Energy Magazine*. July 1999. 16-20.
- * Hall, Nick and John Reed (2001). "Merging Program Theory and Market Theory in the Evaluation Planning Process." *International Energy Program Evaluation Conference*: Salt Lake City, UT. pp. 297-304.
- * Hall, Nick and Jeff Riggert (2003). *Non-Energy Benefits Cross Cutting Report: Year 1 Efforts*. Wisconsin Department of Administration, Division of Energy. January.
- * Hall, Nick and Johna Roth (2003). *Non-Energy Benefits to Implementing Partners from the Wisconsin Focus on Energy Program*. Wisconsin Department of Administration, Division of Energy. October.
- * Hanson, Ralph A and Donna Farrell Siegal (1995). "The Enduring Effects of an Elementary School Energy Education Program." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 493-504.
- Harvey, Andrew (1976). "Estimating regression models with multiplicative heteroscedasticity." *Econometrica* (44): 461-464.
- Hassinger, Edward (1959). "Stages in the Adoption Process." *Rural Sociology* 26: 52-53.
- * Indicates that the information was used in the construction of the Framework.

- * Hastie, Steve, Ralph Prael, Phil Mosenthal, Dimple Gandhi and Barbara Klein (2000). "A Systematic Application of Theory-Based Implementation and Evaluation of Market Transformation Programs." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 6.157-6.168.
- Hausmann, Jerry (1978). "Specification Tests in Econometrics." *Econometrica* 46 (6): 1251-1271.
- Hausmann, Jerry and William Taylor (1981). "Panel Data and Unobservable Individual Effects." *Econometrica* 49: 1377-1398.
- Heckmann, James J (1976). "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models." *Annals of Economic and Social Measurements* Vol 5: 475-492.
- Heckmann, James J (1979). "Sample Selection Bias as a Specification Error." *Econometrica* 47: 153-162.
- Heckmann, James J and Eleanor Singer (1985). *Longitudinal Analysis of Labor Market Data*. Cambridge University Press: Cambridge, MA.
- * Heitfield, Eric, Bruce Mast, Patrice Ignelzi and Peter Rumsey (1996). "A Comparison of Two Net Analysis Methods Using Data from PG&E's Non-Residential New Construction Program." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 6.157-6.168.
- Henninger, Robert H and Michael J Witte (2003). *EnergyPlus Testing with ANSI/ASHRAE Standard 140-2001 (BESTEST)*. United States Department of Energy. May.
- Herman, Patricia, Shel Feldman, Shahanna Samiullah and Kirsten Stacey Mounzih (1997). "Measuring Market Transformation: First You Need a Story ...". *International Energy Program Evaluation Conference*: Chicago, IL. pp. 319-326.
- Heschong Mahone Group (1999). *Daylighting in Schools: An Investigation into the Relationship Between Daylighting and Human Performance*. Pacific Gas & Electric. August.
- * Heschong Mahone Group (1999). *Skylighting and Retail Sale: An Investigation into the Relationship Between Daylighting and Human Performance*. Pacific Gas & Electric. August.
- Heschong Mahone Group (2002). *Time Dependent Valuation (TDV) Economics Methodology*. Pacific Gas & Electric.
- Hewitt, David C (2000). "The Elements of Sustainability." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 6.179-6.190.
- * Hildebrandt, Eric W (1995). "The Value of Improved Information: Using Decision Analysis to Quantify the Value of DSM Impact Evaluation." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 119-126.
- Hill, David, David Nichols and Hannah Sarnow (1999). "The Environmental Benefits of Low-Income Weatherization." *International Energy Program Evaluation Conference*: Denver, CO. pp. 427-434.
- * Hirst, Eric and John Reed (1991). *Handbook of Evaluation of Utility DSM Programs*. Oak Ridge National Laboratory. ORNL/CON-336.

* Indicates that the information was used in the construction of the Framework.

- Inter-Laboratory Working Group (2000). *Scenarios for a Clean Energy Future*. Oak Ridge National Laboratory, Lawrence Berkeley National Laboratory, and National Renewable Energy Laboratory. ORNL/CON-476, LBNL-44029, NREL-TP-29379.
- International Performance Measurement and Verification Protocol (2001). *IPMVP Volume I: Concepts and Options for Determining Energy and Water Savings*.
- Isaac, Stephen and William B Michael (1997). *Handbook in Research and Evaluation: A Collection of Principles, Methods, and Strategies Useful in the Planning, Design, and Evaluation of Studies in Education and the Behavioral Sciences*. 3rd edition. Educational and Industrial Testing Services: San Diego, CA.
- Itron Incorporated (2004). *DEER Update Study*. San Deigo, CA.
- * Ives-Petersen, Dune (1999). "Using the Program Logic Model to Increase the Relevance and Use of Evaluation Findings of Market Transformation Projects." *International Energy Program Evaluation Conference*: Denver, CO. pp. 203-214.
- Jacobs, Pete and Hugh Henderson (2002). *State-of-the-Art Review: Whole Building, Building Envelope and HVAC Component and System Simulation and Design Tools*. Air Conditioning and Refrigeration Institute. ARTI-21CR-605-30010-30020-01.
- Jasso, Guillermina (1985). "Marital Coital Frequency and the Passage of Time: Estimating the Separate Effects of Spouses' Age and Marital Duration, Birth and Marriage Cohort, and Period Influences." *American Sociological Review* 50: 224-241.
- Johnson, Aaron C, Marvin B Johnson and Rueben C Buse (1967). *Econometrics: Basic and Applied*. MacMillan Publishing Company: New York, NY.
- Jordan, Gretchen and Nick Hall (2001). *An Evaluation of Selected Technical Assistance Services Provided by the Federal Energy Management Program: Results of a 1999 Customer Survey*. Sandia National Laboratories.
- Jordan, Gretchen, John Reed and John Mortensen (1997). "Measuring and Managing the Performance of Energy Programs: A Case Study." *National Energy Services Conference*: Washington, DC. pp. 100-111.
- * Jump, David, Devan Johnson and Linda Farinaccio (2000). "A Tool to Help Develop Cost-Effective M&V Plans." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 4.213-4.226.
- Kandel, Adrienne V (1999a). "Evaporative Cooler Rebate Program Cuts Load Significantly, and May Overcome Class Barrier." *International Energy Program Evaluation Conference*: Denver, CO. pp. 801-809.
- Kandel, Adrienne V (1999b). *Instrumented Decomposition: A New Method to Estimate the Net Energy Savings Caused by Efficient Appliance Rebate Programs*. Ph.D. Dissertation.
- Kandel, Adrienne V (1999c). "Instrumented Decomposition: A Two-Stage Method for Estimating Net Savings." *International Energy Program Evaluation Conference*: Denver, CO. pp. 789-800.
- Katipamula, Srinivas (1996). "Modeling Energy Use in Large Commercial Buildings." *ASHRAE Transactions* Vol 102 (Part 2).

* Indicates that the information was used in the construction of the Framework.

- Keating, Kenneth, David Goldstein, Tom Eckman and Tom Miller (1998). "Wheat, Chaff and Conflicting Definitions in Market Transformation." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 7.157-7.169.
- Kema-Xenergy (1992a). *Conservation Potential Study, Vol 1., Results and Methods*. Southern California Gas. Study ID 119.
- Kema-Xenergy (1992b). *SCE Energy Efficiency Potential Study*. Southern California Edison. Study ID 55.
- Kema-Xenergy (2002). *Commercial Sector Energy Efficiency Potential Study*. Pacific Gas & Electric. Study ID 3346.
- Kema-Xenergy (2003a). *California Statewide Commercial Sector Natural Gas Energy Efficiency Potential Study*. Pacific Gas & Electric. Study ID SW061.
- Kema-Xenergy (2003b). *California Statewide Residential Energy Efficiency Potential Study*. Pacific Gas & Electric. Study ID 10023.
- Kema-Xenergy (2003c). *Protocol Development for Demand Response Calculation - Findings and Recommendations*. California Energy Commission. February. Report No. 400-02-017F.
- Kennedy, Peter (1996). *A Guide to Econometrics*. 3rd edition. MIT Press: Cambridge, MA.
- Khawaja, Sami, Douglas Ballou and Karen Schch-McDaniel (1992). "Effects of Weatherization Programs on Low-Income Customer Arrearage." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 7.121-7.127.
- * Kiefer, Kurt (1993). "A Framework for Strategic Evaluation Planning." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 623-628.
- Kirkpatrick, Donald L. (1996). "Techniques for Evaluating Training Programs." *International Journal of Training & Development*. January, 1996.
- Kissock, John Kelley, Jeff S Haberl and David E Claridge (2003). "Inverse Modeling Toolkit: Numerical Algorithms." *ASHRAE Transactions* Vol 109 (Part 2).
- Kmenta, Jan (1971). *Elements of Econometrics*. MacMillan Publishing: New York, NY.
- Knoppel, Helmut and Peder Wolkoff (1992). "Chemical, Microbiological, Health and Comfort Aspects of Indoor Air Quality - State of the Art in SBS." *Chemical and Environmental Science, Vol 4*. Joint Research Centre Ispra: Ispra, Italy.
- Kreider, Jan and Jeff S Haberl (1994). "Predicting Hourly Building Energy Usage: The Great Energy Predictor Shootout." *ASHRAE Transactions* 100 (Part 2).
- Kreider, Jan and Xing An Wang (1991). "Artificial Neural Networks Demonstration for Automated Generation of Energy Use Predictors for Commercial Buildings." *ASHRAE Transactions* 97 (2): 775-779.
- Krueger, Richard A (1994). *Focus Groups: A Practical Guide for Applied Research*. 2nd edition. Sage Publications: Thousand Oaks, CA.
- * Kunkle, Rick and Loren Lutzenhiser (1999). "Beyond Market Transformation: Some Perspectives on Energy Evaluation and Research and the Energy Efficiency Movement." *International Energy Program Evaluation Conference*: Denver, CO. pp. 241-252.

* Indicates that the information was used in the construction of the Framework.

- Kushler, Martin and Ed Vine (2003). *Examining California's Energy Efficiency Policy Response to the 2000/2001 Electricity Crisis: Practical Lessons Learned Regarding Policies, Administration, and Implementation*. American Council for an Energy Efficient Economy. Report Number U033.
- Kushler, Martin, Ed Vine and Dan York (2003). "Using Energy Efficiency to Help Address Electric System Reliability: An Initial Examination of the 2001 Experience." *Energy: The International Journal* 28 (4): 303-317.
- Ledyard, Tom (2003). "Evaluating the Underserved Small C&I Market: Building a Bridge to Implementation." *International Energy Program Evaluation Conference*: Seattle, WA. pp. 627-637.
- Lillard, Lee and Jan Paul Acton (1981). "Seasonal electricity demand and pricing analysis with a variable response model." *The Bell Journal of Economics* 12: 71-92.
- Maddala, GS (1971a). "The Use of Error Components Models in Combining Cross Section with Time Series Data." *Econometrica* 37 (1): 55-72.
- Maddala, GS (1971b). "The Use of Variance Components Models in Pooling Cross Section and Time Series Data." *Econometrica* 39 (2): 341-358.
- Magourik, Jeffrey (1995). "Evaluation of Non-Energy Benefits from the Energy Savings Partners Program." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 155-162.
- Mast, Bruce (1999). "Why Can't We All Just Get Along? A Reconciliation of Economic and Innovation Diffusion Perspectives of Market Transformation." *International Energy Program Evaluation Conference*: Denver, CO. pp. 253-264.
- * Mast, Bruce and Patrice Ignelzi (1996). "Developing Confidence in Your Net-to-Gross Ratio Estimates." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 6.105-6.109.
- Mast, Bruce, Jane Peters, Lori Megdal, Patrice Ignelzi and Noah Horowitz (1998). "Measuring the Market Effects of Utility Programs: Lessons from California." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 7.213-7.223.
- Matthews, Robert (1998). *Fact versus Factions: the Use and Abuse of Subjectivity in Scientific Research*. Aston University.
- McCarthy, Pat M., Glen Galfond, Bijayendra Kumar, and Roger Wright (1985). *Sample Designs for Load Research: The Bootstrap Comparison Procedure*. The Electric Power Research Institute. EA 4232.
- * McLaughlin, John A and Gretchen Jordan (1999). "Logic Models: A Tool for Telling Your Performance Story." *Evaluation and Program Planning* Vol 22 (1): 65-72.
- McLaughlin, John A and Gretchen Jordan (2004). Logic Models: A Tool for Describing Program Theory and Performance. *Handbook of Practical Program Evaluation*. Jossey-Base: San Francisco, CA.
- * McRae, Marjorie (2002). "Sure You Do. Un-Huh: Improving the Accuracy of Self-Reported Efficiency Actions." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 10.189-10.196.

* Indicates that the information was used in the construction of the Framework.

- Meber, Bruce, Shel Feldman, Corey Stone and Elizabeth Tolkin (1997). "Converging On the Effects of Utility Lighting Efficiency Programs." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 327-334.
- Megdal & Associates and ASW Engineering Management Consultants Inc. (1998). *Final Report: Statewide Study of the Retention of Measures Installed Under the Direct Assistance Program (DAP)*. Southern California Edison. Study ID 975.
- Megdal & Associates and ASW Engineering Management Consultants Inc. (1999). *Final Report: Measure Retention Study -- 1994 & 1995 Weatherization Programs (RWRI)*. San Diego Gas & Electric. Study ID 957.
- Megdal, Lori (1998). "Integrating Perspectives from Alternative Disciplines to Understand Market Transformation Policy in Energy Markets." *International Association of Energy Economist Conference*: Quebec City, Quebec Canada. pp. 417-424.
- * Megdal, Lori, Mark C Coleman, Jennifer Ellefsen, Lawrence Pakenas, Helen Kim and Scott Albert (2001). "He Did It! He Did It! Providing Evidence for Causality." *National Energy Services Conference*: Jupiter, FL. pp. 140-150.
- Megdal, Lori, Richard Flood, Beatrice Mayo and Tim Holmes (2001). "When Business Analysis Tools Need to Accompany Program Theory Evaluation within Energy Efficiency Market Transformation Efforts." *American Evaluation Association Meeting*.
- Megdal, Lori, Glenn Haynes and Hasan Rammaha (1993a). "Estimating Takeback (Comfort Increase) for a Low-Income, Loan Program, and a Single Family Rebate Program." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 574-579.
- Megdal, Lori, Glenn Haynes and Hasan Rammaha (1993b). *Loan and Whole House Rebates: Program Evaluation*. City of Austin, Environmental and Conservation Services Department.
- * Megdal, Lori, Dune Ives-Petersen and Andy Ekman (2000). "Local Government Associations as Agents of Change." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 9.287-9.298.
- Megdal, Lori, Allen Lee, Todd Board, Betsy Wilkins and Mary O'Drain (1999). "Using Diffusion and Communications Theory to Expand Market Barrier Examination in MT Measurement." *National Energy Services Conference*: Tucson, AZ. pp. 584-595.
- * Megdal, Lori, Eric Paquette and Jerry Greer (1995a). "The Changing Economy as Part of DSM Impact Evaluations: Evidence from a Large C&I Retrofit Program Evaluation." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 325-331.
- * Megdal, Lori, Eric Paquette and Jerry Greer (1995b). "The Importance of Using Analysis of Covariance (ANCOVA), Diagnostics, and Corrections withing Billing Analysis for Large C&I Customers." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 433-440.

* Indicates that the information was used in the construction of the Framework.

- Megdal, Lori and Hasan Rammaha (1992). "The Development of a Local Energy Efficiency Economic Impact Model for Use in Integrated Resource Planning." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 8.127-8.133.
- * Megdal, Lori, Richard Spellman and Bruce Johnson (1999). "Methods & Measurement Issues for a DSM Evaluation Versus a MT Market Assessment and Baseline Study." *International Energy Program Evaluation Conference*: Denver, CO. pp. 17-27.
- Miller, Terence C. (1990). *Time Series Techniques for Economists*. Cambridge University Press: Cambridge, MA.
- Minium, Edward W and Robert B Clarke (1982). *Elements of Statistical Reasoning*. John Wiley & Sons: New York, NY.
- Morgan, David L (1997). *Focus Groups as Qualitative Research*. Qualitative Research Methods Series No. 16. Sage Publications: Thousand Oaks, CA.
- Mundlak, Yair (1978). "On the Pooling of Time Series and Cross Section Data." *Econometrica* 46 (1): 69-85.
- Nevin, Rick and Gregory Watson (1998). "Evidence of Rational Market Valuation for Home Energy Efficiency." *The Appraisal Journal*. October, 1998: 401-409.
- Nexant and Lawrence Berkeley National Laboratory (2002). *Detailed Guidelines for FEMP M&V Option A*. Federal Energy Management Program.
- * O'Drain, Mary and Tim Caulfield (1994). "Assessing Persistence: Experiences Documenting Savings Persistence Under the California Protocols." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 2.165-2.176.
- * O'Meara, Kevin and Jim Flanagan (1994). "Evaluating Educational Effects in Pacific Gas and Electric's Energy Savings Plan." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 8.151-8.156.
- Ozog, Michael, Ronald Davis and Dorothy Conant (1995a). "Billing Data Analysis of the C&I Sector: Application of Monthly Panel Models." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 447-450.
- Ozog, Michael, Ronald Davis, Don Waldman and Dorothy Conant (1995b). "Model Specification and Treatment of Outliers in the Evaluation of a Commercial Lighting Program." *Energy Services Journal* 1 (1): 55-65.
- * PA Consulting Group Inc. (2003). *Standardized Methods for Free Ridership and Spillover Evaluation - Task 5 Final Report (Revised)*. National Grid, NSTAR Electric, Northeast Utilities, Unitil, Cape Light Compact. June 16.
- Pacific Gas & Electric (2001). Lighting Fixture Demands.
<http://www.pge.com/docs/pdfs/biz/rebates/spc_contracts/2001_manuals_forms/pp_b_ref_table2.pdf>
- * Paquette, Eric (1996). "Why Discrete-Continuous Billing Models Mis-Estimate Net Savings of DSM Programs." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 6.129-6.132.

* Indicates that the information was used in the construction of the Framework.

- Parti, Michael and Cynthia Parti (1980). "The Total and Appliance-Specific Conditional Demand for Electricity in the Household Sector." *Bell Journal of Economics* Vol. 11 (No. 1): 309-321.
- * Peters, Jane (2002). "Effects of Using WebTV Graphics on the Accuracy of Self-Reports." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 10.237-10.246.
- Peters, Jane, Bruce Mast, Patrice Ignelzi and Lori Megdal (1998). *Market Effects Summary Study, Volume 1*. California Demand Side Measurement Advisory Committee. December 15.
- * Peters, Jane, Bruce Mast, Patrice Ignelzi and Lori Megdal (1998). "Measuring Market Transformation: The 1997/1998 California Market Effects Studies." *National Energy Services Conference*: Orlando, FL. pp. 121-128.
- * Peters, Jane, Marjorie McRae, Lyn Morander and Doug O'Brien (2000). "Detecting Behavioral Change from a Visit to a Children's Museum Energy Conservation Exhibit." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 8.281-8.292.
- Phillips, Jack J (1997). *Handbook of Training Evaluation and Measurement Methods*. Butterworth-Heinemann: Woburn, MA.
- Phillips, Jack J and Ron Drew Stone (2002). *How to Measure Training Results: A Practical Guide to Tracking the Six Key Indicators*. McGraw-Hill: New York, NY.
- Phillips, Patricia Pulliam (2002). *The Bottomline on ROI: Basics, Benefits, & Barriers to Measuring Training & Performance Improvement*. CEP Press: Atlanta, GA.
- Pigg, Scott (1994). *An Evaluation of Iowa's Low-Income Weatherization Efforts*. The Wisconsin Energy Conservation Corporation. August.
- Pigg, Scott and Michael Blasnik (1993). "Dealing with Outliers in Impact Evaluations Based on Billing Data." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 188-198.
- Pindyck, Robert S and Daniel L Rubinfeld (1981). *Econometric Models and Economic Forecasts*. McGraw-Hill: New York, NY.
- Pollard, William E (1986). *Bayesian Statistics for Evaluation Research: An Introduction. Contemporary Evaluation Research*. Sage Publications: Beverly Hills, CA.
- Prahl, Ralph and Scott Pigg (1997). "Do the Market Effects of Utility Energy Efficiency Programs Last? Evidence From Wisconsin." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 523-531.
- Presser, Stanley, Jennifer M Rothgeb, Michael P Couper, Judith Lessler, Elizabeth Martin, Jean Martin and Eleanor Singer (2004). *Methods for Testing and Evaluating Survey Questionnaires*. Wiley: Hoboken, NJ.
- Proctor Engineering Group (1996). *Statewide Measure Performance Study: An Assessment of Technical Degradation Rates, Final Report*. CADMAC. April. Study ID 2023.
- Proctor Engineering Group (1998a). *Negative Technical Degradation Factors Supplement to Persistence Studies, Final Report*. Study ID 2031.

* Indicates that the information was used in the construction of the Framework.

- Proctor Engineering Group (1998b). *Statewide Measure Performance Study #2: An Assessment of Technical Degradation Factors, Final Report*. CADMAC. May 14. Study ID 2027.
- Proctor Engineering Group (1999a). *Persistence #3A: An Assessment of Technical Degradation Factors: Commercial Air Conditioners and Energy Management Systems, Final Report*. CADMAC. February 23. Study ID 2028.
- Proctor Engineering Group (1999b). *Summary Report of Persistence Studies: Assessments of Technical Degradation Factors*. CADMAC. February 23. Study ID 2030.
- Proctor, John (1991). "The Development of a Field Furnace Efficiency Test: A More Accurate Prediction of Seasonal Efficiency." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 49-58.
- Public Policy Forum (2003). Six Lessons from the Radwanski Affair.
<http://www.ppforum.ca/ow/ow_zussman_radwanski.htm>
- Quantec (1999). *Fan Speed Reduction in Pneumatic Conveying Systems in Secondary Wood Products Industry*. Northwest Energy Efficiency Alliance. December. Report #99-045.
- Quantec (2002). *ENERGY STAR® Windows, No. 5, Market Progress Evaluation Report*. Northwest Energy Efficiency Alliance. January. Report E01-096.
- Quantum Consulting Inc. (2001). *Statewide LED Traffic Signal Saturation Study*. Study ID 2098.
- Quantum Consulting Inc. (2004). *National Best Practices Study*. California Measurement Advisory Council (CALMAC).
- Quantum Consulting Inc. and Megdal & Associates (2001). *Fourth Year Retention Study for PG&E's 1996 & 1997 Residential AEI Program Refrigeration Technology*. March. Study ID 373 1R1.
- * Randazzo, Katherine Van Dusen, Richard Ridge, Kenneth Train and Leon Clarke (1996). "How Many Mills Ratios Does it Take to Estimate Net Savings?" *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 6.133-6.140.
- Rasmussin, Tami, Kathleen Gaffney and Rob Rubin (2003). "Addressing Program Attribution in the Wake of the California Energy Crisis." *International Energy Program Evaluation Conference*: Seattle, WA. pp. 443-453.
- Reed, John and Nick Hall (1998). *Energy Center Market Effects Study*. Pacific Gas & Electric. May, 1998.
- * Reed, John, Mary O'Drain and Jim Chace (1999). "Transforming Markets Through Education and Information: A Study of the Pacific Energy Center." *International Energy Program Evaluation Conference*: Denver, CO. pp. 841-855.
- Regional Economic Research, Inc. (1991). *Direct Assistance Market Saturation Study*. Study ID 146.
- Reid, Lisa, Michael Baker, Richard Ridge and Bing Tso (1997). "A Combined Engineering and Decision-Analysis Methodology for Evaluating Spillover and Free-Ridership in PG&E's 1995 Industrial Energy Efficiency Program." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 397-404.

* Indicates that the information was used in the construction of the Framework.

- * Richardson, Valerie and Lisa Skumatz (2000). "Measure Retention in Residential New Construction." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 1.235-1.246.
- * Ridge, Richard (1997). "Errors in Variables: A Close Encounter of the Third Kind." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 479-487.
- * Ridge, Richard, Kathleen McElroy and Rob Rubin (2001). "Testing the Causal Linkage Between Training of Sales Personnel in Retail Lighting and Appliance Stores and Changes in Market Share of ENERGY STAR® Qualifying Equipment." *International Energy Program Evaluation Conference*: Salt Lake City, UT. pp. 251-261.
- * Ridge, Richard, Daniel Violette and Donald R. Dohrmann (1994). *An Evaluation of Statistical and Engineering Models for Estimating Gross Energy Impacts*. California Demand Side Management Advisory Committee: The Subcommittee on Modeling Standards for End Use Consumption and Load Impact Models. June.
- Ridge, Richard, Daniel Violette and Donald R. Dohrmann (1998). Appendix J: Quality Assurance Guidelines for Statistical and Engineering and Self-Report Methods for Estimating DSM Program Impacts. *CADMAC Evaluation Protocols*. CADMAC.
- * Riggert, Jeff, Nick Hall, Rick Morgan and Kathy Schroder (1999). "Cinergy's Home Energy House Call (HEHC) Program: An Information Program That Changes People's Lives." *International Energy Program Evaluation Conference*: Denver, CO. pp. 813-827.
- RLW Analytics (1996). *Impact Evaluation of PG&E & SCE's 1994 Non-Residential New Construction Programs*. Pacific Gas & Electric. Study ID 522.1.
- RLW Analytics (1997). *Detroit Edison's 1995 Residential Energy Management Program Arrearage Analysis*. Detroit Edison. December.
- RLW Analytics (1998). *SCE Non-Residential New Construction Persistence Study*. Southern California Edison. Study ID 554.
- RLW Analytics (2000). *Statewide Residential Lighting and Appliances Saturation Study*. Study ID 2094.
- RLW Analytics (2001). *Two Microsoft Access97(TM) Databases and Documentation Relating to Non-Residential New Construction Baseline Study*. CALMAC. Available at <www.calmac.org>.
- RLW Analytics (2003). *Final Report - 1999-2001 Building Efficiency Assessment (BEA) Study: An Evaluation of the Savings By Design Program*. Southern California Edison. Study ID 10029.
- RLW Analytics, Architectural Energy Corporation and ASW Engineering Management Consultants Inc. (1999). *Southern California Edison 1998 Non-Residential New Construction Evaluation*. Southern California Edison. Study ID 572.
- * Robinson, David, David Cohen and Bruce True (1996). "Measure Lifetime Derived from a Field Study of Age at Replacement." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 3.131-3.136.
- Rogers, Everett (1995). *Diffusion of Innovations*. The Free Press: New York, NY.

* Indicates that the information was used in the construction of the Framework.

- * Rogers, Patricia, Timothy A Hacsí, Anthony Petrosino and Tracy A Huebner (2000a). "Program Theory in Evaluation: Causal Models in Program Theory Evaluation." *New Directions for Evaluation* Fall (87): 47-55.
- * Rogers, Patricia, Timothy A Hacsí, Anthony Petrosino and Tracy A Huebner (2000b). "Program Theory in Evaluation: Challenges and Opportunities." *New Directions for Evaluation* Fall (87).
- * Rogers, Patricia, Timothy A Hacsí, Anthony Petrosino and Tracy A Huebner (2000c). "Program Theory in Evaluation: Practice, Promise, and Problems." *New Directions for Evaluation*.
- Rosen, Emanuel (2000). *The Anatomy of Buzz: How to Create Word-of-Mouth Marketing*. Double-Day: New York, NY.
- Rossi, Peter, Howard Freeman and Mark Lipsey (1999). *Program Evaluation: A Systematic Approach*. 6th Edition. Sage Publications: Thousand Oaks, CA.
- Rubin, Herbert J (1983). *Applied Social Research*. Bell & Howell Company: Columbus, OH.
- Rufo, Mike (1993). "DSM Resource Planning the Next Generation: Building the Foundation Through Evaluation." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 162-168.
- Rufo, Mike and Fred Coito (2002). *California's Secret Energy Surplus: The Potentials for Energy Efficiency*. The Energy Foundation. September 23.
- * Samiullah, Shahanna, Miriam Goldberg, Edmund Kademan and Kenneth Train (1996). "Bells, Whistles, and Common Sense: Billing Analysis of a Residential HVAC Rebate Program." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 6.141-6.149.
- * Samiullah, Shahanna, David Hugerford and Adrienne V Kandel (2002). "Do Central Air Conditioner Rebates Encourage Adoption of Air Conditioning?" *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 8.253-8.263.
- San Diego Gas & Electric Company (1991). *Commercial and Industrial Lighting Retrofit Program: Base Equipment Saturation and Operating Hours by Building Type*. Study ID 165.
- San Diego Gas & Electric Company (1998a). *1994 & 1995 Residential Appliance Efficiency Incentives: Compact Fluorescent Lights: Fourth Year Retention Evaluation*. March. Study ID 921.
- San Diego Gas & Electric Company (1998b). *1994 & 1995 Residential Appliance Efficiency Incentives: Refrigerators: Fourth Year Retention Evaluation*. March. Study ID 915.
- Sanderson, Ian (2002). "Evaluation, Policy Learning and Evidence-Based Policy Making." *Public Administration* Vol 80 (No. 1): 1-22.
- Sarndal, Carl-Eric, Bengt Swensson and Jan Wretman (1992). *Model Assisted Survey Sampling*. Springer-Verlag: New York, NY.
- SBW Consulting Inc. and Ridge & Associates (1999). *Pacific Gas & Electric Company PY94 Non-Residential New Construction Retention Study*. Pacific Gas & Electric Company. February 1. Study ID 323R1.

* Indicates that the information was used in the construction of the Framework.

- Schiffman, Dean A (1994). "A Monte Carlo Based Comparison of Techniques for Measuring the Energy Impacts of Demand-Side Management Programs." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 7.213-7.222.
- Schiffman, Dean A and Robert F Engle (1993). *Appendix Z Simulation Study: Comparison of Alternative Methods for Measuring the Gross and Net Energy Impacts of Demand-Side Management Programs (with Addendum)*. San Diego Gas & Electric.
- Schutte, Jeremy and Daniel Violette (1994). "The Treatment of Outliers and Influential Observations in Regression-Based Impact Evaluation." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 8.171-8.176.
- Scriven, Michael (1967a). *AERA Monograph Series on Curriculum Evaluation*. Vol. 1. Rand McNally: Chicago, IL.
- Scriven, Michael (1967b). The Methodology of Evaluation. *Perspectives of Curriculum Evaluation*. R. W. Tyler and R. M. Gagne. Rand McNally: Chicago, IL.
- * Sebold, Frederick D, Alan Fields, Lisa Skumatz, Shel Feldman, Miriam Goldberg, Kenneth Keating and Jane Peters (2001). *A Framework for Planning and Assessing Publicly Funded Energy Efficiency*. Pacific Gas & Electric. March 1. Study PG&E-SW040.
- * Seiden, Ken and Helen Platis (1999). "Freerider and Freedriver Effects from a High-Efficiency Gas Furnace Program." *International Energy Program Evaluation Conference*: Denver, CO. pp. 227-238.
- Shel Feldman Management Consulting, Research Into Action and Xenergy (2001). *The Residential Clothes Washer Initiative: A Case Study of the Contributions of a Collaborative Effort to Transform a Market*. Consortium for Energy Efficiency. June.
- Smith, Peter R, Paul A DeCotis and Karl S Michael (1998). "Linking Market-Based Energy Efficiency Programs to Economic Growth, Sustainable Development and Climate Change Objectives." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 9.185-9.195.
- * Sonnenblick, Richard and Joe Eto (1995). "Calculating the Uncertainty in Estimates of DSM Program Cost-Effectiveness." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 759-768.
- * Spellman, Richard, Bruce Johnson, Lori Megdal and Shel Feldman (2000). "Measuring Market Transformation Progress & the Binomial Test: Recent Experience at Boston Gas Company." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 6.367-6.378.
- Spitler, Jeffrey D, Daniel E Fisher and David C Zietlow (1989). "A Primer on the Use of Influence Coefficients in Building Simulation." *Building Simulation*: Vancouver, BC. pp. 299-304.
- Stout, Jennifer and Steven Scott (2003). *Just Enough Air: Efficient Pneumatic Conveying. Concluding Memorandum*. Northwest Energy Efficiency Alliance. June 28.
- Stram, Daniel and Margaret Fels (1986). "The Applicability of PRISM to Electric Heating and Cooling." *Energy and Buildings* Vol 9 (1,2): 101-110.
- * Indicates that the information was used in the construction of the Framework.

- Sumi, David, Jeff Erickson and Jim Mapp (2002a). "Wisconsin's Public Benefits Approach to Quantifying Environmental Benefits: Creating Different Emissions Factors for Peak/Off-Peak Energy Savings." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 9.353-9.363.
- Sumi, David, Paul Oblander and Ellen Schneider (1993). "A Comparison of Model Specifications in a Billing Data Analysis of Impacts from a Commercial and Industrial Rebate Program." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 256-264.
- Sumi, David, Bryan Ward, Glen Weisbord and Mike Sherman (2002b). "Quantifying Economic and Environmental Benefits." *National Energy Services Conference*: Jupiter, FL. pp. 7.149-7.163.
- TecMRKT Works (2001). California Low Income Public Purpose Test.
- * Tolkin, Betty and Glenn Reed (1993). "Free-Ridership Estimation in the New Construction DSM Market." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 787-791.
- Train, Kenneth (1994). "Estimation of Net Savings from Energy Conservation Programs." *Energy* Vol 19 (No. 4): 423-441.
- * Train, Kenneth, Susan Buller, Bruce Mast, Kirtida Parikh and Eric Paquette (1994). "Estimation of Net Savings for Rebate Programs: A Three-Option Nested Logit Approach." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 7.239-7.247.
- Train, Kenneth, Joseph Herriges and Robert Windle (1985). "Statistically Adjusted Engineering (SAE) Models of End-Use Load Curves." *Energy: The International Journal* Vol 10 (No. 10): 1103-1111.
- Train, Kenneth and Patrice Ignelzi (1987). "The Economic Value of Energy-Saving Investments by Commercial and Industrial Firms." *Energy: The International Journal* Vol 12 (No. 7): 543-553.
- Trochim, William M (2001). *The Research Methods Knowledge Base*. 2nd Edition. Atomic Dog Publishing: Cincinnati, OH.
- USDOE (2003). USDOE Building Energy Software Tools Directory.
<http://www.eere.energy.gov/buildings/tools_directory/software/micropas.htm>
- USEPA (1995a). *Conservation and Verification Protocols, Version 2.0*. U.S. Environmental Protection Agency. 403/B-95-012.
- USEPA (1995b). *Conservation Verification Protocols: A Guidance Document for Electric Utilities Affected by the Acid Rain Program*.
- Vine, Ed (2003). "Opportunities for promoting energy efficiency in buildings as an air quality compliance approach." *Energy* Vol 28: 319-341.
- Vine, Ed, Gregory Kats, Jayant Sathaye and Hemant Joshi (2003). "International Greenhouse Gas Trading Programs: A Discussion of Measurement and Accounting Issues." *Energy Policy* 31 (3): 211-224.
- Vine, Ed and Jayant Sathaye (1999). *Guidelines for the Monitoring, Evaluation, Reporting, Verification and Certification of Energy-Efficient Projects for Climate Change Mitigation*. Lawrence Berkeley National Laboratory. March.

* Indicates that the information was used in the construction of the Framework.

- Vine, Ed and Jayant Sathaye (2000). The Monitoring, Evaluation, Reporting, Verification, and Certification of Energy-Efficiency Projects. *Mitigation and Adaptation Strategies for Global Change*. Kluwer Publishing: Dordrecht, Netherlands. 189-216.
- Violette, Daniel (1991). "Impact Evaluation Accuracy and the Incorporation of Prior Information." *International Energy Program Evaluation Conference*: Chicago, IL. pp. 86-92.
- * Violette, Daniel (1995). *Evaluation, Verification, and Performance Measurement of Energy Efficiency Programs*. International Energy Agency.
- Violette, Daniel, Richard Brakken, Andy Schon and Jerry Green (1993). "Statistically Adjusted Engineering Estimates: What Can the Evaluation Analyst Do About the Engineering Side of the Analysis?" *International Energy Program Evaluation Conference*: Chicago, IL. pp. 656-663.
- * Weiss, Carol H (1998). *Evaluation: Methods for Studying Programs and Policies*. Prentice Hall: Upper Saddle River, NJ.
- Weiss, Carol H (2000). "Which Links in Which Theories Shall We Evaluate?" *Program Theory in Evaluation* Fall (No 87).
- Wilcox, Bruce, Ken Nittler and John Proctor (2001). *Split System Space Cooling Refrigerant Charge and Airflow Measurement*. California Energy Commission. Publication Number P 400-01-014.
- * Woods, James and Sami Khawaja (2000). "Multi-Attribute Valuation for Cost Effective Evaluation of Market Transformation and Other Hard to Quantify Programs." *American Council for an Energy Efficient Economy Summer Study*: Asilomar, CA. pp. 4.227-4.238.
- Worrell, Ernst, John Laitner, Michael Ruth and Hodayah Finman (2001). *Productivity Benefits of Industrial Energy Efficiency Measures*. Lawrence Berkeley National Laboratory. December.
- Worthen, Blaine R, James R Sanders and Jody L Fitzpatrick (1997). *Program Evaluation: Alternative Approaches and Practical Guidelines*. 2nd Edition. Longman: New York, NY.
- Wright, Roger (1983). "Finite Population Sampling with Multivariate Auxiliary Information." *Journal of the American Statistical Association* Vol. 78: 879-884.
- Wright, Roger and David Jacobson (1993). "A Methodology for Integration of Evaluation Studies." *National Demand-Side Management Conference*: Miami Beach, FL. pp.
- * Xenergy (1996). *Net Savings Estimation: An Analysis of Regression and Discrete Choice Approaches*. CADMAC Subcommittee on Base Efficiency.
- Xenergy, ADM Associates, VACom Technologies and Partnership for Resource Conservation (2001). *2001 DEER (Database for Energy Efficiency Resources) Update Study*. California Energy Commission. Study ID 3001.
- Xenergy and Cambridge Systematics Inc. (1993). *Net-to-Gross Ratios for PG&E's CIA Rebate Program*. Pacific Gas & Electric.

* Indicates that the information was used in the construction of the Framework..

APPENDIX C: Guidelines for Evaluation Planning

This Appendix provides some example guidelines for planning and conducting evaluations. These guidelines are provided as a starting point for the evaluation planning process, but evaluation planners are also free to use alternative approaches when submitting evaluation plans. The guidelines may be particularly useful for program implementers with little evaluation experience. The Appendix first describes methods to classify programs according to program type and size. Then the Appendix provides guidelines on selecting specific evaluation activities to meet the evaluation objectives, determining evaluation priorities, and select methods for conducting evaluations.

The information in this appendix is based on several assumptions:

1. Evaluation goals and program policies and procedures are as presented in the CPUC Energy Efficiency Policy Manual (EPPM), version 2 (2003).
2. Program size distributions are similar to the programs conducted during the 2002-2003 program cycle.

As program goals, policies, and procedures evolve, and as program evaluation objectives change, these guidelines will need to be modified to match those changes.

Program Classification

This Framework is one among several potential considerations for the development of evaluation approaches that meet the CPUC's evaluation policy requirements and the associated evaluation efforts. Development of the considerations beyond the Framework is outside of the scope of this project. The function of the Framework is to provide guidance to determine what aspects of the different energy efficiency programs should be evaluated and to assist the evaluation planning process. The guidance provided in the Framework is structured to inform decisions concerning when evaluations are to be conducted, what types of evaluations to conduct, and what methods should be considered for those evaluation efforts. In addition, the Framework provides guidance on the reporting of evaluation results and the discussion (within the reports) on issues related to the accuracy and reliability of the evaluation approach and the results.

There are number of program types discussed in the Framework that have a direct effect on the type of evaluation efforts to be conducted (such as resource acquisition, market transformation, information and education). Each of these program types has a set of decision criteria leading to different evaluation decisions. The highest-level criteria in the Framework is that programs are classified within these broad categories of program type, and that these categories feed into different sets of evaluation decisions. For example, resource acquisition programs have the highest evaluation priority of identifying net program/technology energy impacts (see Chapter 6 of the Framework); market transformation programs have the highest evaluation priority of identifying net

market effects leading to longer-term energy savings (see Chapter 10); and information and education programs have the highest evaluation priority of identifying the effects related to their specific program goals (see Chapter 9). The second highest priority for each of these types of programs is conducting detailed process evaluations (see Chapter 8) focusing on how a program can be changed to better accomplish their primary operational goals. For programs that rely on the value of the program's non-energy effects to achieve their energy impact goals, there may also be a desire to conduct a non-energy effects evaluation (see Chapter 11). Within each of these chapters (6-11), there are discussions of other considerations that may impact the evaluation planning process or the timing of the evaluation. These discussions focus on how the evaluation history of a program, the maturation level of a program, the need for fast feedback information, and other issues can influence the evaluation planning process.

Other information about the program can also be useful in helping decide what efforts to undertake when planning an evaluation. This appendix provides an example of different program classification schemes and approaches for using these schemes to help select the most appropriate impact evaluation methodologies.

A proposed program type classification scheme is presented in Table C.1 below. Each program administrator could assist the process of identifying the appropriate evaluation for their program by classifying their program into program types based on the specific attributes that apply to their program.

Table C.1: Sample Program Classification Scheme

Program Attribute	Description	Applies: (Yes or No)
Program size (based on expected energy impacts)	Small	
	Medium	
	Large	
Program strategies	Audits	
	Codes and standards	
	Commissioning/Operations and Maintenance	
	Design assistance	
	Direct installation	
	Education, training, and information	
	Financing	
	Market transformation	
	Rebate - customized	
	Rebate - prescriptive	
	Performance contracting	
	Upstream	
	Other	
Market segments	Agricultural	
	Commercial	
	Industrial	
	Residential	
	Cross-Cutting	
Market event	New construction	
	Remodel/Renovation	
	Retrofit	
End use/measure groups	Appliances	
	Comprehensive	
	Envelope	
	Food service	
	HVAC	
	Lighting	
	Motor	
	Process	
	Refrigeration	
	Water heating	
	Water pumping/treatment	
	Other	

The proposed classification scheme is a subset of the scheme presented in Chapter 1 of the CPUC Energy Efficiency Policy Manual (EPPM), Version 2,⁴⁰⁵ with modifications to support the implementation of the Framework. A key consideration in the development and use of any classification scheme should include an assessment of how that scheme would apply across different types of evaluation and non-evaluation efforts beyond those

⁴⁰⁵ (* CPUC 2003).

described in this Framework (e.g. load forecasting, efficiency potential studies, DEER update, etc.).

Program Size

The size distribution in terms of net electricity savings of the programs approved for the 2002-2003 program cycle, including utility and third party programs, is shown in Figure C.1. As is evident from this figure, the top three programs are responsible for almost 50% of the total projected net portfolio energy savings.

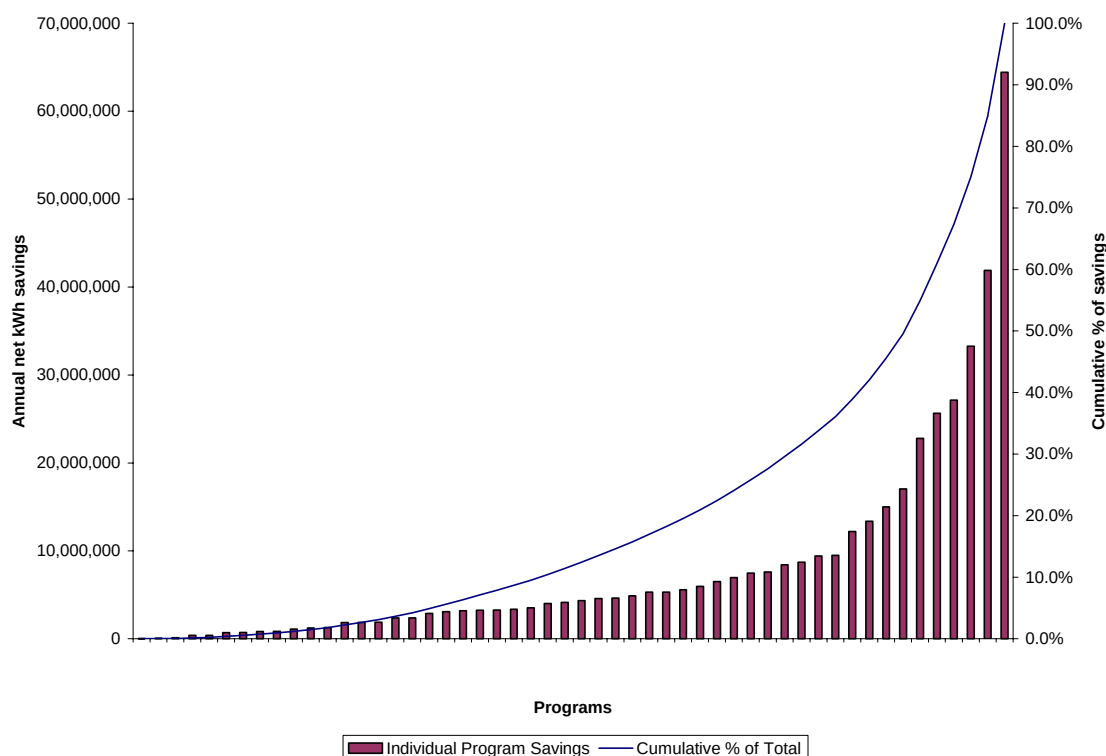


Figure C.1: Program Size Distribution based on Annual kWh Savings

Programs with smaller than 4 GWh net savings per year make up about 10% of the annual net savings expected, but represent nearly 50% of the total count of programs claiming kWh savings.

The size distribution of programs claiming gas savings approved for the 2002-2003 program cycle is shown in Figure C.2.

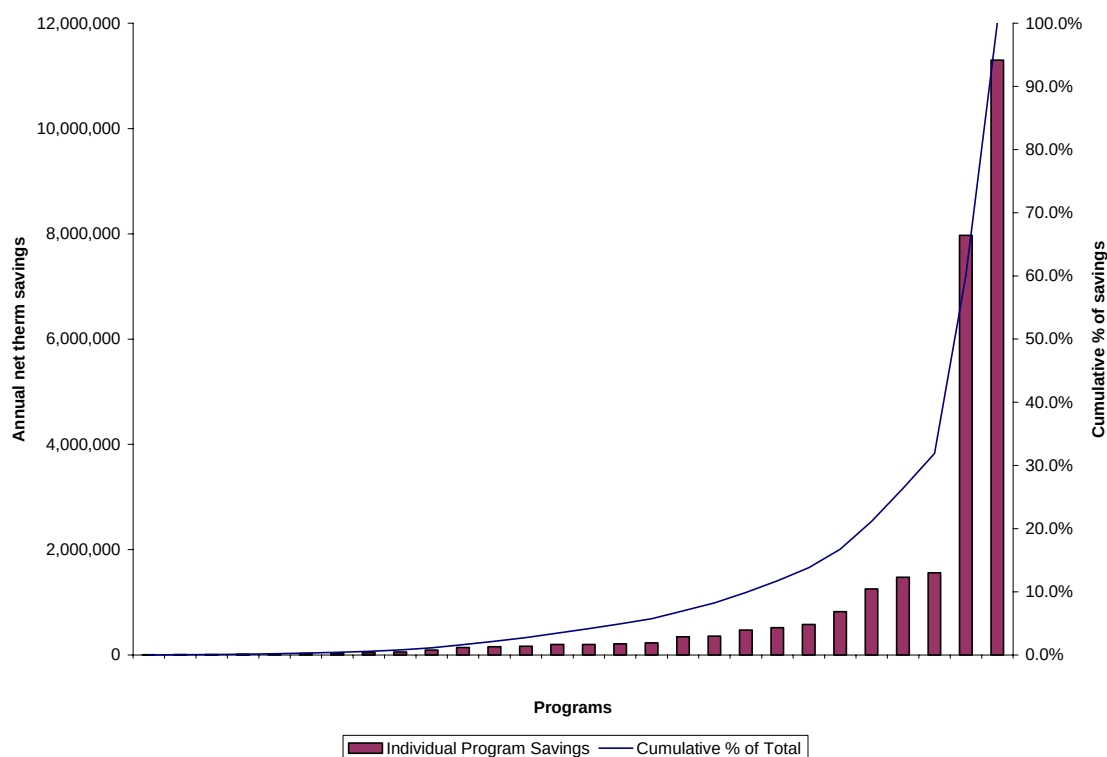


Figure C.2: Program Size Distribution Based on Annual Gas Savings

The two largest programs make up about two-thirds of the expected gas savings. Programs less than 500,000 therms per year of savings make up about 10% of the total portfolio savings, but represent 70% of the total count of programs claiming gas savings. Future program sizes may vary from those offered during this particular cycle, but it is likely that the expected savings from program to program in any program cycle will not be evenly distributed.

A program size classification is presented in Table C.2 below. Programs are classified into bins according to the fraction of the total expected first year savings represented by each bin.⁴⁰⁶ This classification strategy is applied to the portfolio of program presented in the 2002-2003 program cycle. Future portfolios may require a different size classification scheme, based on the size of programs offered and the overall objectives of the portfolio.

⁴⁰⁶ For example, the sum of the savings from all small programs is 5% of the portfolio total, the sum of the savings from all medium sized programs is 45% of the portfolio total, and the sum of the savings of all large programs is the remaining 50% of the total portfolio savings.

Table C.2: Program Size Classification

Size Classification	Electricity			Gas		
	Percent of Total Portfolio Savings	Estimated Program Savings	Number of Programs	Percent of Total Portfolio Savings	Estimated Program Savings	Number of Programs
Small	Up to 5%	Up to 3 GWh/yr	23	Up to 5%	Up to 200 ktherm/yr	16
Medium	5% - 50%	3 to 20 GWh/yr	28	5% - 15%	200 to 1,500 ktherm/yr	10
Large	50%-100%	20+ GWh/yr	6	15%-100%	1,500+ ktherm/yr	2

Program Strategies

Program design strategies referenced in the classification scheme are described below.

Audits. Audit programs involve the inspection of a home, building or industrial process by an expert who makes recommendations on strategies to reduce customer energy consumption. Data relating to the characteristics of the building or process may be collected by an energy auditor during an on-site inspection or supplied by the customer.

Codes and standards. Codes and standards programs provide technical, financial and/or market information and testimony to the California Energy Commission (CEC) for the purpose of enhancing building energy and appliance efficiency standards.

Commissioning/operations and maintenance. Building commissioning provides documented confirmation that building systems as constructed function in accordance with the intent of the building designers, and satisfy the owner's operational needs. Commissioning programs can provide both technical assistance and incentives to support commissioning activities. Commissioning carried out in existing facilities is generally called "Retro-commissioning." Operations and Maintenance programs provide information, engineering analysis and/or incentives to improve the building operations and maintenance practices.

Design assistance. Design assistance programs provide design and analysis services to architects and engineers responsible for the design of new residential and commercial buildings. Such services are structured to influence the design of the building to make it more energy efficient. Design strategies may also emphasize non-energy benefits such as improved indoor air quality, comfort, and lighting quality as a package of features, structured to meet the goals of the building owner or occupant.

Direct installation. Direct installation programs provide free energy efficiency measures for qualified customers. These are generally delivered to the customer and installed without charge. Measures generally distributed in direct install programs include low-cost measures such as compact fluorescent lamps, low-flow shower heads and faucet

aerators; but may also include comprehensive lighting, weatherization or HVAC system tune-up services.

Information and education. Information and education programs can provide a wide range of activities designed to inform or educate a customer or customer group. Generally these range from in-depth, one-on-one, on-site or centrally located classroom style instruction in topics related to energy efficiency, to programs that target information to specific types of customers, to general information provided to a wide range of customers, to short inexpensive public service announcements on FCC approved communication frequencies.

Information and education programs can also develop curriculums and/or provide or prepare presentations for elementary or secondary schools and colleges. These can range from handouts and brochures to supplement a science curriculum, all the way to interactive educational media and projects where the installation of measures at a school is part of the learning process with the students (an education program that complements or interacts with a resource acquisition program).

Information and education programs may also focus on a particular media using specifically targeted audiences or specifically targeted messages. They can include public service announcements; radio, television, or newspaper advertising; cooperative advertising; community events at ball games, schools, fairs, etc.; work with community organizations; trade show events; web-based efforts (including banner advertising and location-specific "click-throughs"); and many other avenues of providing information and education.

Programs that have as their primary objective to inform or educate customers about ways to save, manage, or control their energy use can be considered an information or education program.

Financing. Financing programs encourage investments in energy efficiency through offering below market interest rates or terms, gap financing, financial packaging, and/or simplified administrative procedures.

Market transformation. Market transformation programs seek to improve the adoption of energy efficient technologies and practices by permanently reducing or eliminating market barriers to technology adoption. They seek to "transform" the market to change the market structure, operation, or decision-making in such a way to yield greater opportunity and likelihood for more energy efficient equipment and practices being adopted and sustained. Often these programs are distinguished from resource acquisition programs in that their primary gains in energy and demand savings are not obtained directly, but through market changes that lead to the increased adoption of the desired equipment or practices without direct participant-focused program intervention. (See Chapter 10 on Market Transformation Program Evaluation and the *2001 Framework Study* for additional information on these types of programs and how they are different from resource acquisition programs.)

Rebate – customized. Customized rebate programs provide financial incentives based on an analysis of the customer’s existing equipment and an agreement on the specific products to be installed. The rebate amount is generally tied to the expected energy savings calculated on a customer-by-customer basis.

Rebate – prescriptive. Prescriptive rebate programs provide a prescribed financial incentive per unit of efficient technology installed or per unit of efficiency improvement for a prescribed list of individual products.

Performance contracting. Performance contracting programs consist of contracts between a program implementer and associated contractors (and/or customers) to deliver energy savings from the installation of energy efficiency measures at a customer facility. The measures are installed in exchange for a payment stream based on a portion of verified energy savings. Energy savings are generally verified through a predefined set of Measurement and Verification (M&V) protocols.

Upstream programs. Upstream programs provide information and/or financial assistance to entities involved in the delivery chain of high-efficiency products at the retail, wholesale, or manufacturing level. They can be part of a resource acquisition effort or a market transformation initiative. Incentives are paid to persons other than the eventual end user. Examples include incentives to promote the stocking of high efficiency equipment at the wholesale level, rebates offered to contractors to offset the incremental cost of efficient appliances or equipment, or programs promoting the design and manufacture of efficient products.

Other. This category covers programs not otherwise mentioned in this section.

Market Segments

Programs are generally targeted at customers within specific markets. The program classification scheme uses the following market segment definitions.

Residential. Residential customers are defined as existing single family residences, multi-family dwellings (whether master-metered or individually metered), and customers within facilities that are essentially residential but used for commercial purposes, including but not limited to time shares and vacation homes.

Commercial (and institutional). Customers occupying facilities used for business, commercial and institutional purposes.

Agricultural. Customers engaged in enterprises defined as agricultural by the U.S. Department of Commerce.

Industrial. Customers engaged in enterprises defined as industrial by the U.S. Department of Commerce.

Crosscutting. Programs offered to multiple markets are classified as crosscutting.

Market Event

Programs are designed to engage customers at various market events, including:

Retrofit/early replacement. Retrofit refers to energy efficiency activities undertaken in existing buildings or facilities to promote early replacement of existing inefficient equipment with efficient equipment.

Remodel/renovation. Modifications to the characteristics of an existing building or facility, or the energy-using equipment installed within, involving (1) construction that involves complete removal, redesign, and replacement of the energy consuming systems of a building or process, (2) projects that require design and selection of new systems based upon the needs of new or modified space function(s), and (3) major tenant improvements that add new load. Renovation refers to remodels involving changes to the building shell.

Replacement. Normal replacement of equipment, either as a result of an emergency such as equipment failure, or as part of a broader planned event intended to change equipment near the end of its service life but prior to failure.

New construction. Residential and non-residential buildings that have been newly built or have added major additions subject to Title 24, the California building energy efficiency standards, including (1) new building projects wherein no structure or site footprint presently exist (“greenfield”); (2) addition or expansion of an existing building or site footprint; or (3) addition of new load, as in the example of an existing site adding a new process.

End Use or Measure Groups

Programs may be designed to focus on one or more end uses or measures, as described below.

Appliances. Household appliances such as refrigerators, clothes washers and dryers, room air conditioners, etc. Although these measures are generally promoted to residential customers, they may be applied to any market segment.

Building envelope. Improvements to the building shell, including exterior roofs, walls, windows, air leakage sealing and weatherization.

Food service. Efficient technology for retail preparation of food for sale in restaurants and groceries. Measures generally include improved cook line equipment and commercial kitchen ventilation systems. Food processing programs addressing industrial processing of produce, livestock, dairy products, etc., for wholesale distribution should be classified as an industrial process measure, as described below.

HVAC. Equipment designed to provide heating, ventilation and/or air conditioning. Includes furnaces, central air conditioners, ventilation systems, chillers, boilers, energy management control systems, etc. Efficient motors and motor controls applied to HVAC system fans, pumps, chillers, cooling towers, etc., are included in this category.

Lighting/daylighting. Equipment, control systems, and architectural features designed to provide illumination in buildings. Includes efficient lamps, ballasts, lighting fixtures (luminaries), lighting controls (such as occupancy sensors, dimming controls, timers) and fenestration (such as windows, clerestories, skylights) designed to admit natural light for space illumination.

Motors. Energy efficient motors and controls for non-HVAC drive power applications.

Process. Improvements in industrial processes that reduce the energy consumption per unit of production output. This could include processing equipment, process layout, processing design, and input and output process management (including material waste-stream management).

Refrigeration. Improvements to refrigeration systems applied to grocery stores, restaurants, food processing, refrigerated warehouses and industrial applications. Includes efficient compressors, oversized condensers, close approach evaporators, systems controls, etc. Shell and lighting improvements to refrigerated warehouses are generally classified as refrigeration, not building envelope measures.

Water heating. Efficient equipment applied to service or potable water heating, including water heaters, boilers, water heater controllers, water heat tank and pipe insulation.

Water pumping/treatment. Equipment applied to the pumping, storage and treatment of drinking water, wastewater, and agricultural irrigation.

Comprehensive. Multiple end uses or measure groups applied to building or facility in the context of an integrated or systems approach

Other. Measures or end uses not otherwise listed.

CPUC Evaluation Objectives

The CPUC Energy Efficiency Policy Manual, Version 2⁴⁰⁷ lists the following objectives for program evaluations:

1. Measure energy and peak savings;
2. Measure cost-effectiveness;
3. Provide upfront market assessment and baseline analysis;
4. Provide ongoing feedback and guidance to the program administrator;
5. Measure indicators of effectiveness and testing program theory and approach;
6. Assess the overall levels of performance and success;
7. Inform decisions regarding compensation and final payments; and
8. Help assess the continuing need for the program.

The role of the principal evaluation disciplines (impact, process, market effects, and non-energy effects) in meeting these objectives is described below.

Measure Energy and Peak Savings

Impact evaluation, supported by measurement and verification is primarily used to provide estimates of resource acquisition program energy and demand savings. Likewise, evaluations that focus on identifying changes in market effects can be used to inform an impact evaluation that estimates longer-term energy savings from market transformation programs.

Measure Cost-Effectiveness

Cost-effectiveness is measured from an estimate of the program costs and benefits, as described in Chapter 14 and the *California Standard Practice Manual*.⁴⁰⁸ Impact evaluation is the primary method for establishing program benefits in terms of energy and demand savings. Non-energy effects, as they relate to the cost-effectiveness calculations also may play a role for programs able and allowed to claim non-energy effects in their program accomplishments. However, the non-energy effects are not to be included in an assessment of the cost-effectiveness associated with acquiring energy resources. Program costs and the accuracy of the reporting systems and operational processes may be investigated during process evaluations that inform cost-effectiveness assessments.

⁴⁰⁷ (* CPUC 2003).

⁴⁰⁸ (* California State Governor's Office 2001).

Provide Upfront Market Assessment and Baseline Analysis

Market assessments and market baseline analysis are generally conducted through market transformation program evaluations. These broad, market-wide evaluations are used as a reference to establish program-specific baselines. Each program is responsible for determining the appropriate baseline conditions for the program's operation and evaluation. Impact evaluations generally rely on a clear definition of the program baseline, though baseline analyses are generally not conducted as a component of impact evaluation.

Provide Ongoing Feedback and Guidance to the Program Administrator

Process evaluations and Measurement and Verification (M&V) studies are conducted to inform the impact assessment, but are also conducted to provide feedback on the program's operations, measure installation quality, in-field measure performance, installation contractor quality, public relations, etc. Using the process evaluation for feedback to the program implementer is discussed in more detail in Chapter 8.

Measure Indicators of Effectiveness and Testing Program Theory and Approach

This goal is primarily addressed during process evaluations, with input from the impact and M&V studies. The Chapter 4 discussion on program theory can help provide some background and guidance for understanding and using the program theory for this purpose.

Assess the Overall Levels of Performance and Success

Overall levels of performance and success are measured on many levels and include input from impact, process, market, and non-energy effects evaluations as appropriate. This evaluation goal is met through the combination of all of the evaluation efforts associated with a specific type of program and suggests that a comprehensive evaluation approach is needed for each program. The Framework is constructed so that this goal will be met through the combination of program-specific evaluations.

Inform Decisions Regarding Compensation and Final Payments

At this time, compensation and final payments are based on ex-ante savings estimates and program accomplishments based on numbers of installations completed. M&V and process evaluations are used to verify the measure installation counts for final payment determination.

Help Assess the Continuing Need for the Program

Several information sources can be used to assess the continuing need for a program. These include market potential studies, formal or informal market assessments, documented program accomplishments and accomplishment rates, energy saving goal

attainment and attainment rates, program cost-effectiveness ratios, process evaluation results, program participation rates, program outreach and promotional success and other information can be used to help assess the continuing need for a program. Since multiple programs may operate in a particular market, it may not be possible for any particular program evaluation to judge the effect of the program on the overall market. The decision to continue the program rests with the portfolio administrator and the CPUC rather than the program administrator, but cost-effectiveness, accomplishments, net impacts and process evaluation results are key components to the overall decision. Key information to inform this decision includes:

- Overall accomplishments, in terms of total number of measures installed and customers contacted.
- Estimate of the market penetration and remaining market that could be served by the program, including program enrollment waiting lists.
- Cost-effectiveness of the program, based on ex-post net impacts and verified program costs.
- Measures of customer satisfaction from process evaluations.

The role of each evaluation discipline in meeting the CPUC evaluation goals is shown in Table C.3.

Table C.3: Evaluation Goals and Study Type

CPUC Evaluation Goal	Evaluation Type			
	Impact	Process	Market Effects	Non-Energy Effects
Measure energy and peak savings	●			
Measure cost-effectiveness	●			●
Provide upfront market assessment and baseline analysis	○		●	
Provide ongoing feedback and guidance to the program administrator	●	●		
Measure indicators of effectiveness and testing program theory and approach		●		○
Assess the overall levels of performance and success	●	●		○
Inform decisions regarding compensation and final payments	●		○	
Help assess the continuing need for the program		●	○	

Note: ● indicates primary role; ○ indicates secondary role

Evaluation Priorities

The CPUC evaluation goals direct evaluators to conduct impact and process evaluations of all programs. Non-energy effects evaluations may also be included if the program theory indicates that non-energy effects are a key factor in attaining the program energy and demand impact goals. Market evaluations are generally conducted on a market-wide rather than on an individual program basis, with the exception of market transformation programs. Guidelines for establishing the priorities in evaluation rigor and resource allocation are described below.

Risk to Portfolio Performance

Evaluation resources would yield the greatest overall value by being allocated based upon minimizing uncertainty at the portfolio level. Thus, large programs with uncertain or unproven impact results would receive more resources than small or predictable programs. For example, the uncertainty in the expected savings is a function of a number of factors, including:

- **Measure types promoted.** Evaluation knowledge and user circumstances vary by measure type. Therefore, uncertainty also varies by measure type. Lighting measures in general are fairly well characterized based on the long operating history of lighting efficiency programs in California. HVAC measures are less certain, due to variability in operating hours, control, and building load characteristics. Measures involving building controls, operations and maintenance and/or commissioning are even less certain, due to the possibility of occupant tampering and changes that can defeat these measures.
- **Market penetration and delivery mechanism.** Expected savings are based on the ability of the program to meet measure installation goals. If a particular program is having trouble meeting installation goals due to a flawed program design or inadequate market interest, then the expected savings are at greater risk. Programs that are well designed and well established, and operate in markets that have sufficient demand represent a lower risk to the portfolio.

The risk assessed and minimized based on evaluation results is not only that of the current portfolio, but to some future portfolio of programs. Programs that are currently small, but may contain significant upside potential in a future efficiency portfolio may justify increased evaluation resources to minimize future risks. Therefore, a small program that would appear at first to warrant the use of lower cost evaluation strategy may warrant a more extensive evaluation based on future growth potential.

Prior Evaluation History

The Framework suggests that program implementers examine their program evaluation requirements and construct an evaluation plan with evaluation activities scheduled throughout the program cycle so that evaluation efforts can be scheduled and funded as they are needed. The Framework Umbrella (Chapter 5) generally suggests that program

implementers conduct full program net impact evaluation during each program cycle. However, the scope of the evaluation activities should consider recent evaluation history in order to provide the best use of evaluation resources. Evaluation planners should consider the following questions when developing evaluation plans:

- Is the program new to the market?
- How recent was the last program evaluation?
- Have there been substantial changes since the last evaluation in terms of technologies, procedures, rebate levels, energy costs, economic activity, delivery mechanisms, etc., that may substantially alter the results from the prior evaluation?

New programs should plan to conduct full net impact evaluations during the first program cycle. Continuing programs experiencing stable impact results can propose to skip a program cycle, based on a justification addressing the issues described above. See Chapter 5, Umbrella Framework, for more information concerning when an impact evaluation should be conducted.

Method Selection for Impact Evaluations

The analysis technique applied to the program impact evaluation depends primarily on the following factors:

- Program size
- Market event
- Expected impacts as a fraction of total billing
- Program strategy

Program Size

Impact evaluation resources should be directed at minimizing the uncertainty in the estimate of the efficiency resource at the portfolio level (or minimizing the risk in efficiency investments). This logically will place more evaluation resources and increased rigor on programs with large expected impacts and/or large expected uncertainty.

Market Event

New construction, remodeling and renovation programs generally require engineering analysis, since pre-program billing data are not available.⁴⁰⁹

⁴⁰⁹ However, it may be possible to conduct a billing analysis for a residential new construction program if a non-participant comparison group can be identified and the econometric model includes variables for climate, economic activity, building size and orientation, and program participation. This approach may work best in circumstances with close matching, such as where participants and non-participants are identified within a single large subdivision of homes with similar design.

Expected Impacts as a Fraction of Total Billing

As a general rule, efficiency programs may need to affect the customer's monthly energy consumption by at least 10% for a billing analysis to be feasible. Smaller impacts as a percentage of the total consumption may not be discernable from the random noise in the billing data. For example, the annual energy consumption impact of a CFL replacement program may not be large enough to allow the use of a billing analysis. Commercial comprehensive retrofit programs may provide energy savings on the order of 20% to 30%, which can be detected by the regression models used in a billing analysis.

Program Strategy

The program strategy presents some unique issues to consider when choosing an analysis method for conducting an impact evaluation. Guidelines based on program strategy are presented below.

Audits. Audit programs are generally classified as information only and therefore do not require impact evaluations but do need program effects evaluations. Audit programs that claim energy savings may be evaluated using a billing analysis, provided the energy impacts of the installed measures are sufficiently large relative to the monthly bill. To conduct an engineering analysis, the measures that were ultimately installed as a result of the audit must be identified, and engineering estimates of savings applied to each of the measures. This may prove to be problematic, since the decision to install a measure may have been influenced by multiple programs. For example, some audit programs direct customers to incentive programs covering recommended measures. In these cases, a joint evaluation of the two programs might provide the most useful information and cost-effective evaluation.

Codes and standards. Codes and standards programs generally affect new construction and therefore require an engineering based impact analysis.⁴¹⁰

Commissioning/operations and maintenance. Commissioning and O&M programs are generally targeted at HVAC systems. Calibrated building energy simulation models should be used to estimate impacts of commissioning programs on new construction. Billing analysis can be used for retro commissioning and O&M projects, provided the impact is sufficiently large relative to the billing data.⁴¹¹ Commissioning and O&M programs are especially sensitive to savings degradation over time. A high quality evaluation effort would specify a clear method for quantifying the persistence of savings over time.

⁴¹⁰ The 2005 Title 24 Building Energy Standards contain provisions for existing buildings upon equipment replacement. Impact evaluation of existing building provisions only may be candidates for billing analysis.

⁴¹¹ A retro commissioning program operated by AEC for Southern California Edison averaged 13% energy savings.

Design assistance. Design assistance programs can be classified as information only, and therefore do not require impact evaluations but they do need program effects evaluations. (See Chapter 9 on Information and Education Program Evaluation.) Design assistance programs targeting new construction that are claiming energy savings will use engineering methods. Most design assistance programs take a comprehensive look at the building design. Whole building impacts may be best analyzed using building energy simulation programs.

Direct installation. Direct installation programs are a form of rebate program where the program bears 100% of the measure cost. These programs generally target small impact measures, such as CFL replacements, low-flow showerheads, faucet aerator, etc. Billing data analysis may be difficult due to the modest impact these measures have on overall energy consumption. Matching participants with utility account numbers may also be difficult due to the simplified participation rules for many of these types of programs. Engineering analysis informed by field observations conducted under IPMVP Option A⁴¹² may be the most appropriate for direct install programs.

Information and education. Information and education programs provide only information or educational services, and therefore do not need an impact evaluation. (See Chapter 9 on Information and Education Program Evaluation to make this determination.) However, program effects evaluations are often appropriate and can help to assess the program's justification and determine the program's cost-effectiveness (measured as cost per effect rather than as in the *Standard Practice Manual*⁴¹³ where energy and/or demand savings can be measured).

Financing. Financing programs generally operate in cooperation with information type programs, such as audit or design assistance programs, to help customers finance the energy efficiency measures. In some cases, they work in combination with incentive programs to help finance the customer portion of the costs of a measure or subsidize the interest rate. "Double-dipping," where savings resulting from the installation of a particular measure are claimed by more than one program, must be avoided if energy savings are claimed as part of the incentive program. Financing programs that claim energy savings may be evaluated using billing analysis provided the energy impacts of the installed measures are sufficiently large. To conduct an engineering analysis, the measures that were ultimately installed as a result of the program must be identified, and engineering estimates of savings applied to each of the measures. In order for this to be done efficiently, program applications need to contain information about the specific measures financed and include information about the equipment replaced.

Rebate – customized. Custom rebate programs use calculations prepared by the program implementer to calculate energy saving, rebate amounts, and customer economics. The initial energy savings estimates can be used in combination with a billing analysis, or M&V can be used to verify program calculations in a calibrated engineering model

⁴¹² (International Performance Measurement and Verification Protocol 2001).

⁴¹³ (* California State Governor's Office 2001).

approach. Custom rebates are used in both retrofit and new construction programs; new construction programs often will use calibrated simulation models.

Rebate – prescriptive. Prescriptive rebate programs provide precalculated rebate or cash amounts to a standard list of measures. An engineering based approach informed from field measurements conducted under the IPMVP Option A can be used for smaller programs. Billing analysis is most often applied to larger rebate programs, provided the load impact is sufficiently large.

Performance contracting. Performance contracting programs are generally designed with an integrated set of M&V activities. Impact data can be gathered from the M&V reports, provided the M&V activities meet the evaluation standards. Timing of payments and M&V activities for performance contracting programs is generally different than rebate programs, since the payments are made over time and are based on verified energy savings.

Upstream programs. Upstream programs offer incentives, information, and/or training to manufacturers, distributors, and retailers to build, stock and/or promote energy efficient products. Program data on customer locations where efficient products are installed may not be available. These programs can be resource acquisition (being tied to specific installations of acquired savings) or market transformation programs. Chapter 10 on Market Transformation Program Evaluation provides information on how to assess the accomplishments of those upstream efforts that are market transformation initiatives.

End Use/Measure Groups

Guidelines for impact evaluation strategies as they apply to specific end use or measure groups are summarized below.

Appliances. Load impacts from appliance programs often do not affect billing data by more than 10%, thus billing analysis may not be successful for appliance efficiency programs. (Exceptions to this may include refrigerator programs, particularly in homes without electric heat or central air conditioning or homes that use electricity to heat their domestic water.) Engineering analysis, supported by M&V is a preferred evaluation approach for these types of measures. The partially measured, retrofit isolation approach from the International Performance Measurement and Verification Protocol (IPMVP Option A) is appropriate for small programs; medium and large programs may want to employ the retrofit isolation approach (IPMVP Option B).

Comprehensive. Comprehensive programs involving multiple, interacting measures generally require a billing analysis or building energy simulation approach. Retrofit programs where measure impacts are expected to exceed 10% of the billing data can utilize a billing analysis approach to get a program-level estimate of impacts. Engineering estimates informing a billing analysis (SAE approach) may benefit from field measurements taken under the IPMVP Option A approach. The IPMVP Whole Facility M&V approach (Option C) may be used for programs with few participants.

New construction programs generally require a building energy simulation model approach, using a program such as MicroPas or DOE-2. Model calibration should be conducted according to the IPMVP Calibrated Simulation (Option D) approach. Industrial programs may require the use of a specialized computer model of the process addressed by the program.

Envelope. Programs addressing building envelope improvements may be evaluated using a billing analysis approach provided the expected impacts exceed 10% of the billing data. Residential buildings receiving comprehensive weatherization services generally fall into this category. The impact of building shell improvements as a fraction of commercial building consumption is generally less than residential building consumption, requiring an engineering based analysis. Small programs may rely on engineering equations informed by M&V studies under IPMVP Option A. Larger programs may find greatest value from using a building energy simulation approach, with the models calibrated under IPMVP Option D.

Food service. Programs involving food service efficiency improvements in commercial buildings may use a billing analysis, provided the impacts are 10% or more of the baseline billing data. In commercial buildings such as restaurants and groceries, this may be the case for estimating the program impacts on gas consumption, but will not likely be the case for electricity, unless the customer uses electricity for their major cook line equipment or is in a small facility without central air-conditioning. Engineering analysis, informed by field measurements taken under IPMVP Options A or B is most appropriate for commercial food service equipment efficiency upgrades. Programs involving commercial kitchen ventilation and HVAC issues should be evaluated using a building energy simulation program, calibrated under IPMVP Option D.

HVAC. Programs addressing HVAC system improvements expected to impact energy billing by more than 10% may use a billing analysis. This may be the case in residential buildings, but will not likely be the case in commercial buildings. Engineering estimates informing a billing analysis using an SAE approach may benefit from field measurements taken under the IPMVP Option A. Small programs may use engineering analysis informed by field measurements taken under IPMVP Option A. For medium and large programs a building energy simulation program may be more appropriated with those calibrated to field measurements under the IPMVP Option D.

Lighting. Lighting programs conducted for residential customers will not likely achieve enough savings to allow a billing analysis. Engineering analysis informed by field measurements conducted under IPMVP Option A for small programs or IPMVP Option B for larger programs is the preferred approach. Commercial lighting programs may be evaluated using a billing analysis technique if the impacts are expected to be greater than 10% of the billed consumption. Otherwise, engineering analysis informed by field measurements taken under IPMVP Options A or B is appropriate.

Motors. Due to the relatively small impacts of motor efficiency programs relative to total billed consumption, motor efficiency programs probably need to be evaluated using

engineering analysis. Small programs can use field measurements under IPMVP Option A, while larger programs may be justified using IPMVP Option B to inform the engineering analysis.

Process. If the impacts are sufficiently large, a whole-building billing analysis based on IPMVP Option C may be conducted. Otherwise, engineering analysis informed by field measurements under IPMVP Option A for small programs or IPMVP Option B for larger programs may be the most appropriate methods. Specialized engineering models developed for a specific process can also be used. These models would then be calibrated according to IPMVP Option D.

Refrigeration. Refrigeration is a major end use in grocery stores, thus comprehensive retrofit of refrigeration systems may be analyzed with a billing analysis. Engineering estimates informing a billing analysis (SAE approach) may benefit from field measurements taken under the IPMVP Option A. Otherwise, an engineering-based approach could alternatively (or jointly) be used. Grocery store refrigeration systems generally have significant interactions with building HVAC systems, thus engineering approaches based on building energy simulation modeling are the more appropriate engineering approach. Small programs may use engineering analyses that account for HVAC interactions, informed by field measurements conducted under IPMVP Option A. An approach appropriate for larger programs would be to use building energy simulation modeling calibrated under IPMVP Option D.

Refrigerated warehouse programs may utilize a billing analysis approach under IPMVP Option C if the impacts are sufficiently large. Otherwise, simple engineering equations informed by field measurements taken under IPMVP Options A or B should be used for small programs. A building energy simulation approach using a simulation model suitable for refrigerated warehouses, calibrated according to IPMVP Option D would be more appropriate for larger programs.

Water heating. Water heating programs may sufficiently impact gas consumption in climates with moderate heating loads to allow for a billing analysis. Engineering estimates informing a billing analysis (SAE approach) may benefit from field measurements taken under the IPMVP Option A. Otherwise, engineering analysis informed by field measurements under IPMVP Option A for small programs or IPMVP Option B for larger programs might be the appropriate options.

Water pumping/treatment. If the impacts are sufficiently large, a billing analysis based on IPMVP Option C may be conducted. Otherwise, engineering analysis informed by field measurements under IPMVP Option A for small programs or IPMVP Option B for larger programs might be appropriate. Specialized engineering models developed for a specific waterworks or wastewater treatment processes can also be used. These models are then higher quality efforts if be calibrated according to IPMVP Option D.

A summary of the recommendations for impact evaluation and M&V approach is shown in Table C.4.

Table C.4: Example Guidelines for Impact Evaluation Methodology by Program Type and Size

Program Size	Market Segments	Market Event	End use/ measure groups	Impact as % of billing	Gross Impact Method	M&V Option	Net-to-gross method
All	All	All	Appliances	< 10%	Engineering analysis	A	Survey-based
All	All	All	Appliances	< 10%	Engineering analysis	B	Survey-based
All	Residential, Commercial	All	Comprehensive	< 10%	Building energy simulation	D	Survey-based
All	Residential, Commercial	New construction/ Remodel/Renovation	Comprehensive	> 10%	Building energy simulation	D	Survey-based
All	Residential, Commercial	Retrofit	Comprehensive	> 10%	Billing analysis	A optional	Econometric
All	Agricultural, Industrial	All	Comprehensive	< 10%	Engineering analysis	A	Survey-based
All	Agricultural, Industrial	New construction/ Remodel/Renovation	Comprehensive	> 10%	Engineering analysis	A	Survey-based
All	Agricultural, Industrial	Retrofit	Comprehensive	> 10%	Billing analysis	C	Survey-based
All	All	New construction/ Remodel/Renovation	Envelope	> 10%	Building energy simulation	D	Survey-based
Small	All	Retrofit	Envelope	< 10%	Engineering analysis	A	Survey-based
Medium, Large	All	Retrofit	Envelope	< 10%	Building energy simulation	D	Survey-based
Small	All	Retrofit	Envelope	> 10%	Billing analysis	A optional	Econometric
Medium, Large	All	Retrofit	Envelope	> 10%	Billing analysis	A	
Small	Commercial	All	Food service	< 10%	Engineering analysis	A	Survey-based
Medium, Large	Commercial	All	Food service	< 10%	Building energy simulation	D	Survey-based
All	Commercial	Retrofit	Food service	> 10%	Billing analysis	A	Econometric
All	Commercial	New construction/ Remodel/Renovation	Food service	> 10%	Building energy simulation	D	Survey-based
All	All	Retrofit	HVAC	> 10%	Billing analysis	A	Econometric
All	All	New construction/ Remodel/Renovation	HVAC	> 10%	Building energy simulation	D	Survey-based
Small	All	All	HVAC	< 10%	Engineering analysis	A	Survey-based
Medium, Large	All	All	HVAC	< 10%	Building energy simulation	D	Survey-based

Table C.4: Continued

Program Size	Market Segments	Market Event	End use/ measure groups	Impact as % of billing	Gross Impact Method	M&V Option	Net-to-gross method
Small	All	All	Lighting	< 10%	Engineering analysis	A	Survey-based
Medium, Large	All	All	Lighting	< 10%	Engineering analysis	B	Survey-based
Small	All	Retrofit	Lighting	> 10%	Billing analysis	A	Econometric
Medium, Large	All	Retrofit	Lighting	> 10%	Billing analysis	B	Econometric
Small	All	New construction/ Remodel/Renovation	Lighting	> 10%	Engineering analysis	A	Survey-based
Medium, Large	All	New construction/ Remodel/Renovation	Lighting	> 10%	Engineering analysis	B	Survey-based
Small	All	All	Motors	< 10%	Engineering analysis	A	Survey-based
Medium, Large	All	All	Motors	< 10%	Engineering analysis	B	Survey-based
Small	Agricultural, Commercial, Industrial	All	Process	< 10%	Engineering analysis	A	Survey-based
Medium, Large	Agricultural, Commercial, Industrial	All	Process	< 10%	Engineering analysis	B	Survey-based
All	Agricultural, Commercial, Industrial	Retrofit	Process	> 10%	Billing analysis	C	Survey-based
All	Agricultural, Commercial, Industrial	New construction/ Remodel/Renovation	Process	> 10%	Engineering analysis	B	Survey-based
All	Commercial	Retrofit	Refrigeration	> 10%	Billing analysis	A	Econometric
All	Commercial	New construction/ Remodel/Renovation	Refrigeration	> 10%	Building energy simulation	D	Survey-based
Small	Commercial	All	Refrigeration	< 10%	Engineering analysis	A	Survey-based
Medium, Large	Commercial	All	Refrigeration	< 10%	Building energy simulation	D	Survey-based
Small	Agricultural, Industrial	All	Refrigeration	< 10%	Engineering analysis	A or B	Survey-based
Medium, Large	Agricultural, Industrial	All	Refrigeration	< 10%	Building energy simulation	D	Survey-based
All	Agricultural, Industrial	Retrofit	Refrigeration	> 10%	Billing analysis	C	Survey-based
All	Agricultural, Industrial	New construction/ Remodel/Renovation	Refrigeration	> 10%	Building energy simulation	D	Survey-based

Sampling Criteria

Establishing evaluation priorities and methods is an exercise in balancing the available evaluation budgets with meeting the evaluation goals for each program without placing too much burden on programs with limited resources. Applying evaluation techniques and choosing sample sizes that are appropriate given the program size, budget, and risk to the portfolio can maintain this balance. The evaluation approach guidelines above direct more rigorous evaluation efforts toward larger programs. This approach would also direct more robust sampling strategies at larger programs.

A high quality evaluation plan would examine the program characteristics, evaluation resources, and the potential tradeoffs between spending resources on increased sample size versus using alternative evaluation methods, and/or approaches and the potential effects on precision and minimization/mitigation of bias. Then the selected methods and sample sizes being recommended would be selected for the most reliable overall evaluation that meets the evaluation goals (to include tradeoffs between different types of evaluations that would provide value at this time and the various methods and samples needed for each of these).

Data Requirements

Impacts at the measure level are desirable, but not required under the Energy Efficiency Policy Manual, Version 2 (the guiding document at the time of the drafting of the Framework). It is recommended that, if at all possible, program net impacts for energy and demand be reported by costing period. All quality impact evaluations include an estimate and discussion of the uncertainty, which includes sampling error and uncertainty in engineering calculations and field measurements as discussed in Chapter 12 and 13 in this Framework. Potential biases and their influence on program results should be identified and reported. For ease of use, regardless of the error bound level (10%, 15%, 25%, or other), the Framework recommends that all uncertainty calculations be expressed at 90% confidence, to facilitate the calculation of portfolio-level savings and uncertainties.



STATE OF CALIFORNIA

PUBLIC UTILITIES COMMISSION

505 Van Ness Avenue
San Francisco, California 94102

California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals

APRIL 2006

Prepared for the **California Public Utilities Commission**

by **The TecMarket Works Team**

**Under Contract with and Directed by the CPUC's
Energy Division, and with guidance from Joint Staff**

California Energy Efficiency Evaluation Protocols: Technical, Methodological and Reporting Requirements for Evaluation Professionals {a.k.a. Evaluators' Protocols}

Prepared under direction of the **Energy Division**, with the guidance by **Joint Staff**, for the
California Public Utilities Commission

APRIL 2006

Submitted by

Nick Hall, Johna Roth, Carmen Best
TecMarket Works

TecMarket Business Center
165 West Netherwood Road, Second Floor, Suite A
Oregon, WI 53575
NPHall@TecMarket.net
608 835 8855

And sub-contractors

Sharyn Barata

Opinion Dynamics, Irvine, California

Pete Jacobs

BuildingMetrics, Inc., Boulder, Colorado

Ken Keating, Ph.D.

Energy Program Consultant, Portland, Oregon

Steve Kromer

RCx Services, Oakland, California

Lori Megdal, Ph.D.

Megdal & Associates, Acton, Massachusetts

Jane Peters, Ph.D.

Research Into Action, Portland, Oregon

Richard Ridge, Ph.D.

Ridge & Associates, Alameda, California

Francis Trottier

Francis Trottier Consulting, Penn Valley, California

Ed Vine, Ph.D.

Energy Program Consultant, Berkeley, California

Acknowledgements

The authors wish to acknowledge and express our appreciation to the many individuals who contributed to the development of the California Evaluation Protocols. Without the support and assistance of these individuals this effort would not have been possible.

The Joint Staff (California Public Utilities Commission and the California Energy Commission) provided considerable Protocol development guidance and conducted multiple rounds of reviews of all sections of the Protocols. These individuals and their affiliations are the following:

- Ariana Merlino, Project Manager, Energy Division, California Public Utilities Commission
- Mike Messenger, California Energy Commission

Appreciation is also extended to the Administrative Law Judge, Meg Gottstein, who ordered the development of the Protocols and who provided instructive guidance and policy direction along the way.

In addition to the oversight and guidance provided by the above individuals, others within the Energy Division of the California Public Utilities Commission and the California Energy Commission provided valuable contributions and support. For these efforts we thank the following individuals:

- Nancy Jenkins, California Energy Commission
- Tim Drew, Energy Division, California Public Utilities Commission
- Zenaida Tapawan-Conway, Energy Division, California Public Utilities Commission
- Peter Lai, Energy Division, California Public Utilities Commission
- Nora Gatchalian, Energy Division, California Public Utilities Commission
- George Tagnipes, Energy Division, California Public Utilities Commission
- Sylvia Bender, California Energy Commission

We also wish to thank the California investor owned utilities and their program management and evaluation staff who have attended workshops and provided both written and verbal comments during the Protocol development process. And we wish to thank the public representatives who attended the workshops and provided verbal and written comments. All of these combined efforts helped move the development of the Protocols to a successful completion in a very short period of time.

Lastly, we wish to thank the TecMarket Works Protocol Project Team who under direction from the ALJ and the Joint Staff, and with useful comments from the IOUs and the public, took the Protocols from concept to completion under the oversight of Joint Staff.

This team was made up of the following individuals:

- Nick Hall, Johna Roth, Carmen Best, TecMarket Works

- Sharyn Barata, Opinion Dynamics Corporation
- Pete Jacobs, Building Metrics Inc.
- Ken Keating, Ken Keating and Associates
- Steve Kromer, RCx Services
- Lori Megdal, Megdal & Associates
- Jane Peters, and Marjorie McRae, Research Into Action
- Rick Ridge, Richard Ridge and Associates
- Ed Vine, Edward Vine and Associates
- Francis Trottier, Francis Trottier Consulting

Table of Contents

ACKNOWLEDGEMENTS	I
TABLE OF TABLES.....	IX
TABLE OF FIGURES.....	XI
CALIFORNIA ENERGY EFFICIENCY EVALUATION PROTOCOLS: TECHNICAL, METHODOLOGICAL AND REPORTING REQUIREMENTS FOR EVALUATION PROFESSIONALS {A.K.A. EVALUATORS' PROTOCOLS}.....	1
INTRODUCTION	1
HOW THE PROTOCOLS WERE DEVELOPED	5
HOW THE PROTOCOLS WORK TOGETHER	6
HOW THE PROTOCOLS MEET CPUC GOALS	7
USE OF THE EVALUATION RESULTS TO DOCUMENT ENERGY SAVINGS AND DEMAND IMPACTS.....	9
THE EVALUATION IDENTIFICATION AND PLANNING PROCESS	9
EVALUATION RIGOR AND BUDGETS	13
<i>Impact Evaluations.....</i>	<i>13</i>
<i>Process Evaluations</i>	<i>14</i>
<i>Market Effects Evaluations</i>	<i>14</i>
<i>Codes and Standards and Emerging Technology Program Evaluations</i>	<i>14</i>
<i>Evaluation Budgets</i>	<i>15</i>
RECOMMENDATIONS FOR USING THE PROTOCOLS	15
THE DETAILED EVALUATION WORK PLAN	15
CONFIDENTIALITY ISSUES.....	17
CONTACTING THE CUSTOMER.....	18
IMPACT EVALUATION PROTOCOL	19
INTRODUCTION	19
AUDIENCE AND RESPONSIBLE ACTORS.....	21
OVERVIEW OF THE PROTOCOL	21
<i>Protocol Types.....</i>	<i>21</i>
<i>Rigor.....</i>	<i>22</i>
<i>Key Metrics, Inputs and Outputs.....</i>	<i>23</i>
ENERGY AND DEMAND IMPACT PROTOCOLS	25
<i>Gross Energy Impact Protocol.....</i>	<i>25</i>
<i>Gross Demand Impact Protocol.....</i>	<i>32</i>
<i>Participant Net Impact Protocol</i>	<i>36</i>
INDIRECT IMPACT EVALUATION PROTOCOL	40
GUIDANCE ON SKILLS REQUIRED TO CONDUCT IMPACT EVALUATIONS	45
SUMMARY OF PROTOCOL-DRIVEN IMPACT EVALUATION ACTIVITIES	46
MEASUREMENT AND VERIFICATION (M&V) PROTOCOL	49
INTRODUCTION	49
AUDIENCE AND RESPONSIBLE ACTORS.....	50

OVERVIEW OF THE PROTOCOL	50
<i>M&V Framework & Language</i>	51
<i>Relationship of the M&V Protocol to Other Protocols</i>	51
<i>Key Metrics, Inputs and Outputs</i>	52
SITE-SPECIFIC M&V PLAN	54
M&V RIGOR LEVELS	56
MEASURE INSTALLATION VERIFICATION	56
<i>Measure Existence</i>	56
<i>Installation Quality</i>	56
<i>Correct Operation and Potential to Generate Savings</i>	57
M&V PROTOCOL FOR BASIC LEVEL OF RIGOR	57
<i>IPMVP Option</i>	57
<i>Sources of Stipulated Data</i>	58
<i>Baseline Definition</i>	58
<i>Monitoring Strategy and Duration</i>	58
<i>Weather Adjustments</i>	58
M&V PROTOCOL FOR ENHANCED LEVEL OF RIGOR	58
<i>IPMVP Option</i>	59
<i>Sources of Stipulated Data</i>	59
<i>Baseline Definition</i>	60
<i>Monitoring Strategy and Duration</i>	60
<i>Weather Adjustments</i>	60
<i>Calibration Targets</i>	60
<i>Additional Provisions</i>	61
M&V APPROACH EXAMPLES	61
OVERALL RESULTS REPORTING	62
SAMPLING STRATEGIES	63
SKILLS REQUIRED FOR M&V	63
SUMMARY OF PROTOCOL-DRIVEN M&V ACTIVITIES	63
EMERGING TECHNOLOGIES PROTOCOL	65
INTRODUCTION	65
<i>Audience and Responsible Actors</i>	66
<i>Key Metrics, Inputs, and Outputs</i>	66
<i>Evaluation Planning</i>	67
<i>A Sample of Available ETP Evaluation Methods</i>	68
<i>Protocols Requirements</i>	69
<i>Integration of Results</i>	75
<i>Reporting of Results</i>	75
<i>Summary</i>	76
<i>Summary of Protocol-Driven Emerging Technology Evaluation Activities</i>	76
<i>References</i>	77
CODES AND STANDARDS AND COMPLIANCE ENHANCEMENT EVALUATION PROTOCOL	81
INTRODUCTION	81
<i>Audience and Responsible Actors</i>	82
<i>Key Inputs, and Outputs</i>	83
<i>Evaluation Methods</i>	84
<i>Evaluation Planning</i>	84
<i>Technology-Specific Code and Standard Change Theory</i>	85
<i>Evaluation Approach</i>	88

Summary.....	98
Summary of Protocol-Driven Codes and Standards Evaluation Activities	99
CODE COMPLIANCE ENHANCEMENT PROGRAMS	100
<i>Definition of a Code Compliance Enhancement Program</i>	<i>100</i>
<i>What this Protocol is Designed To Do</i>	<i>100</i>
<i>Joint Staff Responsibilities</i>	<i>100</i>
DRAFT EVALUATION PLAN	101
<i>Program Theory Review and Assessment.....</i>	<i>101</i>
<i>Pre-Program Compliance Rate.....</i>	<i>101</i>
<i>Post-Program Compliance Rate.....</i>	<i>102</i>
<i>Adjustment For Naturally Occurring Compliance Change.....</i>	<i>102</i>
<i>Net Program-Induced Compliance Change</i>	<i>102</i>
<i>Assessment of Energy Savings.....</i>	<i>103</i>
<i>Recommendations for Program Changes.....</i>	<i>103</i>
<i>Cost Effectiveness Assessment.....</i>	<i>103</i>
REPORTING OF EVALUATION RESULTS	103
REFERENCES	104
EFFECTIVE USEFUL LIFE EVALUATION PROTOCOL (RETENTION AND DEGRADATION).....	105
INTRODUCTION	105
AUDIENCE AND RESPONSIBLE ACTORS.....	107
OVERVIEW OF THE PROTOCOL	108
<i>Protocol Types.....</i>	<i>108</i>
<i>Rigor.....</i>	<i>109</i>
<i>Key Metrics, Inputs, and Outputs.....</i>	<i>110</i>
RETENTION STUDY PROTOCOL	111
DEGRADATION STUDY PROTOCOL	116
EFFECTIVE USEFUL LIFE ANALYSIS PROTOCOL.....	119
REPORTING REQUIREMENTS	126
STUDY SELECTION AND TIMING	127
GUIDANCE ON SKILLS REQUIRED TO CONDUCT RETENTION, EUL, AND TECHNICAL DEGRADATION EVALUATIONS	128
<i>Summary of Protocol-Driven Impact Evaluation Activities</i>	<i>130</i>
PROCESS EVALUATION PROTOCOL	131
INTRODUCTION	131
AUDIENCE AND RESPONSIBLE ACTORS.....	132
OVERVIEW OF THE PROTOCOL	132
PROCESS EVALUATION PLANNING.....	133
ANNUAL PROCESS EVALUATION PLANNING MEETING	134
RECOMMENDATIONS FOR CHANGE	135
KEY ISSUES AND INFORMATION COVERED	135
PROCESS EVALUATION EFFORTS	136
<i>Program-Specific Process Evaluation Plans</i>	<i>137</i>
<i>Data Collection and Assessment Efforts</i>	<i>137</i>
<i>Conducting Investigative Efforts</i>	<i>138</i>

<i>Independence</i>	140
<i>Selection of Evaluation Contractors</i>	140
SKILLS REQUIRED FOR CONDUCTING PROCESS EVALUATIONS	140
MARKET EFFECTS EVALUATION PROTOCOL.....	143
INTRODUCTION	143
<i>Overview of the Market Effects Protocol</i>	143
<i>The Market Effects Protocol and Other Protocols</i>	146
<i>Key Market Effects Inputs and Outputs</i>	147
AUDIENCE AND RESPONSIBLE ACTORS.....	148
STEPS IN CONDUCTING MARKET EFFECTS EVALUATIONS	148
<i>Scoping Study</i>	149
<i>Collection of Baseline and Longitudinal Indicators</i>	154
<i>Analysis of Market Effects</i>	155
<i>Reporting</i>	158
GUIDANCE ON SKILLS REQUIRED TO CONDUCT MARKET EFFECTS EVALUATIONS.....	159
CONSIDERATIONS FOR CONDUCTING A MARKET EFFECTS EVALUATION.....	159
SUMMARY OF PROTOCOL-DRIVEN MARKET EFFECTS EVALUATION ACTIVITIES.....	160
SAMPLING AND UNCERTAINTY PROTOCOL.....	163
INTRODUCTION	163
PRECISION: GROSS AND NET IMPACT, MEASUREMENT AND VERIFICATION, AND VERIFICATION ACTIVITIES	164
DEVELOPMENT OF THE EVALUATION STUDY WORK PLAN	167
PROCESS EVALUATIONS	169
MARKET EFFECTS.....	170
SYSTEM LEARNING	170
ACCEPTABLE SAMPLING METHODS	171
SKILLS REQUIRED FOR SAMPLING & UNCERTAINTY	171
AUDIENCE AND RESPONSIBLE ACTORS.....	171
KEY METRICS AND INFORMATION COVERED.....	172
<i>Sample Size and Precision</i>	172
<i>Validity and Research Design</i>	172
<i>Accuracy</i>	173
<i>Summary of Sampling and Uncertainty Protocol</i>	175
EVALUATION REPORTING PROTOCOL	177
INTRODUCTION	177
<i>Report Delivery Dates</i>	177
<i>Common Evaluation Reporting Requirements</i>	177
<i>Evaluation Type Specific Reporting Requirements</i>	183
SAMPLE REPORTING TABLES	195
EVALUATION SUPPORT INFORMATION NEEDED FROM ADMINISTRATORS..	205
<i>Storage and Disposal of Customer Information Used in the Evaluation</i>	211
APPENDIX A. MEASURE-LEVEL M&V RESULTS REPORTING REQUIREMENTS	213
APPENDIX B. GLOSSARY OF TERMS	217
APPENDIX C. PERFORMANCE BASIS METRICS	247

APPENDIX D. A PRIMER FOR USING POWER ANALYSIS TO DETERMINE SAMPLE SIZES	249
BASICS OF POWER ANALYSIS AND THE PROTOCOLS	250
EXAMPLE OF VARYING PARAMETERS AND ESTIMATING REQUIRED SAMPLE SIZE FOR SURVIVAL ANALYSIS THROUGH POWER ANALYSIS	251
REFERENCES	254
APPENDIX E. SUMMARY TABLES FOR ALL PROTOCOLS	255
<i>Summary of Protocol-Driven Impact Evaluation Activities</i>	<i>256</i>
<i>Summary of Protocol-Driven M&V Activities.....</i>	<i>262</i>
<i>Emerging Technology</i>	<i>264</i>
<i>Codes and Standards.....</i>	<i>265</i>
<i>Effective Useful Life</i>	<i>267</i>
<i>Summary of Protocol-Driven Market Effects Evaluation Activities</i>	<i>270</i>
<i>Sampling and Uncertainty.....</i>	<i>274</i>

Table of Tables

Table 1. Required Protocols for Gross Energy Evaluation.....	26
Table 2. Required Protocols for Gross Demand Evaluation.....	33
Table 3. Required Protocols for Participant Net Impact Evaluation	36
Table 4. Required Protocols for Indirect Impact Evaluation	41
Table 5. Summary of M&V Protocol for Basic Level of Rigor	57
Table 6. Summary of M&V Protocol for Enhanced Level of Rigor	59
Table 7. Model Calibration Targets.....	60
Table 8. Programs Compliant with ASHRAE Standard 140-2001 (Partial List)	61
Table 9. Example IPMVP Options by Measure Type and Rigor Level	61
Table 10. Sample of Available ETP Evaluation Methods	68
Table 11. Required Protocols for Measure Retention Study	113
Table 12. Required Protocols for Degradation Study	118
Table 13. Required Protocols for EUL Analysis Studies	122
Table 14. Required Protocols for Market Effects Evaluation Scoping Studies	150
Table 15. Required Protocol for Market Theory and Logic Models	151
Table 16. Types of Market Interventions and Associated Possible Indicators	153
Table 17. Required Protocol for Market Effects Evaluation Indicator Studies	155
Table 18. Required Protocol for Preponderance of Evidence Approach to Causal Attribution Estimation	156
Table 19. Required Protocols for Gross Impacts.....	165
Table 20. Required Protocols for Net Impacts.....	166
Table 21. Required Protocols for Measure-level Measurement and Verification	167
Table 22. Required Protocols for Sampling of Measures Within a Site.....	167
Table 23. Required Protocols for Verification.....	167
Table 24. Measure-Level Impact Reporting Requirements.....	213
Table 25. Sample Sizes as a Function of Alpha and Power.....	253

Table of Figures

Figure 1. Operational Overview of How the Protocols Relate to Each Other	7
Figure 2. The Program Evaluation Planning Process for Programs, Program Components and Program-Covered Technologies	12
Figure 3. The Market Effects Evaluation Planning Process	13
Figure 4. Required Protocols for Direct Impact and Indirect Impact Evaluations	23
Figure 5. Potential Alternative Behavioral Impact Paths.....	43
Figure 6. Measurement & Verification Information Flow Diagram.....	53
Figure 7. An Example of How Findings Across the Three Types of Studies Would Work Together for Persistence Evaluations.....	109
Figure 8. Protocols and Rigor Levels for EUL Evaluations	110
Figure 9. Cumulative Logistic Function	123
Figure 10. Logistic Survival Function with EUL=15 and b=0.2	124
Figure 11. Sources of Energy Efficiency Changes in the Market.....	145

California Energy Efficiency Evaluation Protocols: Technical, Methodological and Reporting Requirements for Evaluation Professionals {a.k.a. Evaluators' Protocols}

Introduction

This chapter presents and describes the California Energy Efficiency Evaluation Protocols: Technical, Methodological and Reporting Requirements for Evaluation Professionals (a.k.a. Evaluators' Protocols, referred to hereafter collectively as the Protocols and individually as Protocol) that are designed to meet California's evaluation objectives.

This document is to be used to guide the efforts associated with conducting evaluations of California's energy efficiency programs and program portfolios launched after December 31, 2005. The Protocols are the primary guidance tools policy makers will use to plan and structure evaluation efforts and that staff of the California Public Utilities Commission's Energy Division (CPUC-ED) and the California Energy Commission (CEC) (collectively the Joint Staff), and the portfolio (or program) administrators (Administrators) will use to plan and oversee the completion of evaluation efforts. The Protocols are also the primary guidance documents evaluation contractors will use to design and conduct evaluations for programs implemented after December 31, 2005. This chapter provides an introduction to, and overall guidance for, the use of specific Protocols presented in later chapters of this document.

The Protocols are significantly grounded in the *California Evaluation Framework* of June 2004¹ (*Evaluation Framework*). The Protocols reference the *Evaluation Framework* and other documents that provide examples of applicable methods. The requirements for conducting evaluation studies, however, are always those stated in the Protocols, which take precedence over other evaluation guidance documents, unless otherwise approved or required by the CPUC. That is, these Protocols are the primary evaluation guidance documents for all types of evaluations presented in these Protocols, however this is not to be construed as limiting the ability of the CPUC or the Joint Staff to evaluate items in addition to or beyond those identified in these Protocols or to use evaluation processes and procedures beyond those presented in these Protocols. While these Protocols are the key guiding documents for the program evaluation efforts, the CPUC and the Joint Staff reserve the right to utilize additional methodologies or approach if they better meet the CPUC's evaluation objectives and when it serves to provide reliable evaluation results using the most cost-efficient approaches available. In addition, the Protocols should be considered a "living" document that may need to be updated and revised from time to time as standard evaluation approaches evolve and as Joint Staff and Administrators gain experience using the Protocols. The CPUC will determine when an update is necessary and what process will be used to complete any updates that the agency deems necessary. Protocol users should always confirm that they are referring to the most recently CPUC-approved and -adopted version, which can be found on the CPUC website.

¹ TecMarket Works, *The California Evaluation Framework* (Southern California Edison Company, 2004). The report can be obtained on the CALMAC Web site at: <http://www.calmac.org/search.asp>. Enter "California Evaluation Framework" and download the 500-page reference document as an Adobe .pdf file.

Most of the Protocols are designed to function within an evaluation planning process that focuses on the evaluation needs within a given program cycle. This planning process is described in a other documents adopted by the ALJ and the CPUC, and most directly at part of what are known as the Process Protocols.

The Protocols cover several types of evaluation efforts. The evaluation types covered include the following: direct and indirect impact {including the associated measurement and verification approaches (M&V)}, market effects, emerging technology, codes and standards and process evaluations. In addition, the Protocols provide specific guidelines for conducting effective useful life studies and how evaluation samples should be selected. The primary goal of this document is to specify **minimum** acceptable evaluation approaches and the operational environments in which evaluations are conducted. The primary purpose of the Protocols is to establish a uniform approach for:

- Conducting robust and cost-efficient energy efficiency evaluation studies;
- Documenting ex-post evaluation-confirmed (i.e. realized) energy efficiency program and portfolio effects;
- Supporting the performance bases for judging energy efficiency program and portfolio achievements; and
- Providing data to support energy efficiency program and portfolio cost-effectiveness assessments.

The Protocols may have other uses such as providing support for improving ex-ante energy and demand savings estimates.

This document includes a separate Protocol for each of the following categories:

- Impact Evaluation - Direct and Indirect Effects
- Measurement and Verification
- Process Evaluation
- Market Effects Evaluation
- Codes and Standards Program Evaluation
- Emerging Technology Program Evaluation
- Sampling and Uncertainty Protocol (for use in determining evaluation sampling approaches) Reporting Protocol (to guide evaluation data collection and reporting)
- Effective Useful Life Protocol (used to establish the period over which energy savings can be relied upon)

The Protocols also include information on the type of evaluation-related information and support needed from program administrators and implementers in order to conduct the evaluation efforts. The purpose of each of the listed Protocols is described below.

Impact Evaluation Protocol: The Impact Evaluation Protocol prescribes the minimum allowable methods to meet a specified level of rigor that will be used to measure and document the program or program component impacts achieved as a result of implementing energy efficiency programs and program portfolios. Impact evaluations estimate net changes in

electricity usage, electricity demand, therm usage and/or behavioral impacts that are expected to produce changes in energy use and demand. Impact evaluations are limited to addressing the direct or indirect energy impacts of the program on participants, including participant spillover impacts. However, while the Protocols provide for the assessment of participant spillover, these results are not to be counted toward program or portfolio energy savings goal accomplishments, and as such are to be distinctly and separately identified in any impact reporting.² The impact evaluation studies are also not expected to document program influences on the operations of a market or the program's impacts on non-participants. Program-induced changes that affect non-participants or the way a market operates are addressed in the Market Effects Evaluation Protocol. Results from the impact evaluations will support a cost-effectiveness assessment at the program and portfolio level.

Measurement and Verification (M&V) Protocol: The M&V Protocol is designed to prescribe how field measurements and data collection will be conducted to support impact evaluations, updates to ex-ante measure savings estimates and process evaluations.

Process Evaluation Protocol: The Process Evaluation Protocol is designed to support Administrator (i.e. Investor Owned Utility or IOU) efforts to conduct evaluations that both document program operations and provide the basis for improving the operations or cost-effectiveness of the programs offered within the portfolio.

Market Effects Evaluation Protocol: The Market Effects Evaluation Protocol is designed to guide evaluations conducted to document the various market changes that affect the way energy is used within a market and estimate the energy and demand savings associated with those changes that are induced by sets of program or portfolio interventions in a market.

Codes and Standards Program Evaluation Protocol: The Codes and Standards Program Evaluation Protocol is designed to guide evaluation approaches for codes and standards programs.

Emerging Technology Program Evaluation Protocol: The Emerging Technology Program Evaluation Protocol is designed to guide evaluation approaches for emerging technology programs.

Effective Useful Life Protocol: The Effective Useful Life Protocol is designed to guide evaluation approaches for establishing the effective useful life of program measures, including approaches for evaluating measure retention and technical degradation of measure performance. The effective useful life of a measure is the period of time over which program-induced energy impacts can be relied upon.

² The Protocols prescribe minimum requirements for how to conduct and report evaluations. The Performance Basis Protocol takes precedence with regard to including savings toward program or portfolio goals and performance measurement. The most recent CPUC decision will always take precedence and be used for the interpretation and application of the Protocols.

Sampling and Uncertainty Protocol: The Sampling and Uncertainty Protocol is designed to prescribe the approach for selecting samples and conducting research design and analysis in order to identify, mitigate and minimize bias in support of the Protocols identified above.

Reporting Protocols: The Reporting Protocol prescribes the way in which evaluation reports are to be delivered and the way information is to be presented in those reports.

Evaluation Support Information Needed from Administrators: The Protocol document also includes a chapter on the types of information Administrators shall provide to contractors conducting evaluation studies covered by the Protocols.

The four primary types of Evaluation Protocols that cover the majority of California's program offerings are the Impact, M&V, Market Effects and Process Protocols. These are supported by the Sampling and Uncertainty Protocol. However, there are two types of programs that are different enough in their scope and intended results that they require a separate Evaluation Protocol (Codes and Standards and Emerging Technology). As such, two Evaluation Protocols are directed to a specific type of program (Codes and Standards and Emerging Technology), while the remaining Protocols either operate to establish a minimum set of allowable methods for a specific type of evaluation or in support thereof. Any program, program component or set of programs could be included within each of these types of evaluations. The difference lies not in which programs are eligible for which types of evaluations, but in the purpose of and outputs from each of these evaluation types.

The outputs from an impact (and its associated M&V efforts) evaluation are program or program component net energy, demand or behavioral impacts from program participation. Those from a market effects evaluation are energy and demand impacts created by market changes caused by a program or set of programs. While a process evaluation produces the documentation and assessment of program processes, and recommendations to improve them. A program could easily be included in all three types of evaluations. For example, a single program of great significance with respect to the overall portfolio might be directly assessed using impact and process evaluations and also be included in a market effects evaluation for all programs operating in a given market sector.

While it is important to know what is in these Protocols (above), it is also important to know what is not included in these Protocols. These Protocols do not cover the evaluation or research approaches for the following types of programs, efforts or activities:

- Low-income program evaluations;
- Market research for program design, planning or operations;
- Technical, market or other types of potentials studies;
- Meta-evaluations or comparative studies using evaluation study results;
- Demand response programs;
- Renewable energy programs;
- On-site or distributed generation or combined heat and power programs;
- Green house gas or pollution reduction studies;
- Cost-effectiveness methods, approaches or procedures;

- Forecasting methods, approaches or procedures; and
- Public Interest Energy Research (PIER) evaluation efforts.

While it is expected that the Protocols will need to be updated from time to time, it is also expected that new Protocols may need to be added to this document as the need for different types of information evolves. For example, California may need to establish Protocols for crediting greenhouse gas reductions resulting from the energy efficiency program portfolios or for addressing demand response programs that are currently outside the scope of the Protocols.

How The Protocols Were Developed

The Protocols were developed over two different but overlapping three-month timelines involving a number of activities, including presentations to the public and the receipt of public comments and recommendations. The Impact, M&V, Process, Market Effects, Sampling and Reporting Protocols were developed first, and followed by the development of the Codes and Standard, Emerging Technology, and Effective Useful Life Protocols. All of the Protocols were developed using the following approach:

1. The consulting team that the CPUC-ED contracted to develop the Protocols (TecMarket Team) assembled and reviewed comments from previous Protocol and performance basis workshops and comments received during the development of the *Evaluation Framework*;
2. Using the *Evaluation Framework*, previous comments and discussions with the Joint Staff, draft concept Protocol outlines were developed. These concepts were then discussed within a series of meetings with the Joint Staff leading to the development of a set of draft concept Protocols;
3. The draft concept Protocols were presented in public workshops. During the workshops, the attending public was requested to comment on the draft concept Protocols. These comments were recorded and summarized in workshop notes and used to inform Protocol development. At this time, the draft concept Protocols were also placed on the CPUC website for additional public review. An announcement was sent to the CPUC Energy Efficiency service lists advising the public of the workshops and the draft concept Protocol postings. These efforts allowed both attendees and non-attendees of the workshop to review the draft concept Protocols and provide comments;
4. Following the workshop, the TecMarket Team collected comments from both workshop attendees and non-attendees. These comments were distributed to and reviewed by the Joint Staff and the TecMarket Team and used to guide the draft Protocol development efforts;
5. The TecMarket Team developed a set of draft Protocols under the direction of CPUC-ED staff and in consultation with the Joint Staff. The draft Protocols were provided to the Joint Staff for review and comment in order to identify concerns and issues that needed to be addressed in the final draft Protocols. Upon reviewing the draft Protocols, the Joint Staff requested modifications to the Protocols;

6. The TecMarket Team modified the draft Protocols consistent with direction provided by CPUC-ED staff, in consultation with Joint Staff, and provided them to the CPUC-ED project manager for final review and editing;
7. The CPUC-ED project manager submitted the draft Protocols to the ALJ for review and acceptance;
8. The ALJ, in consultation with the CPUC-ED project manager and Joint Staff, reviewed and accepted the final Protocols.
9. The ALJ adopted these Protocols via a Ruling, per the authority delegated her by the CPUC.

In addition to the process outlined above, the first set of Protocols developed (Impact, M&V, Process, Market Effects, Sampling and Reporting) went through an additional round of public review and comment, Joint Staff review and commentary, and CPUC-ED project manager approval and editing process before they were provided in final form to the ALJ for review and acceptance.

How the Protocols Work Together

The Protocols are designed to support the need for public accountability and oversight, the need for program improvements (especially cost-effectiveness improvements) and the documentation of effects from publicly funded or rate-payer funded energy efficiency programs provided in California. The individual Protocols are designed to work together to achieve these goals.

The Impact Evaluation Protocol is meant to guide the design of evaluations that provide reliable ex-post participant-focused net program impacts. These net impacts include peak demand (kilowatts (kW) of electricity), energy (kilowatt-hours (kWh) of electricity and therms of natural gas) and behavioral impacts. The Protocol is focused such that program level impacts can be summed to estimate impacts at the Administrator portfolio level. The Protocol also allows for impact estimates at the program component delivery level (e.g., direct install, participant rebate and information distribution) or at the technology level (e.g., CFLs, motors, HVAC tune-up and refrigerators) when the specific evaluation is meant to acquire these metrics.

The Impact Evaluation Protocol does not operate in isolation from the other Protocols. The M&V Protocol supports impact evaluations and can often serve in a feedback or support role for process evaluations if coordinated to do so. Similarly, the Sampling and Uncertainty Protocol is designed to support impact evaluations, as well as M&V, and process and market effects evaluations by assuring that the sampling designs provide unbiased estimates based on the information needs associated with each evaluation effort. Finally, the Reporting Protocol is designed to support all of the evaluation activities by detailing the information that must be reported for each type of evaluation. The entire evaluation process is facilitated by the additional identification of the information Administrators need to provide the evaluation contractors.

The Protocols, and the evaluations conducted under them, support several efforts. For example, many of the evaluation results, especially the impact evaluation results and the verification aspects of the M&V Protocol, are designed to support program performance assessment,

including the performance-based metrics associated with ex-post energy savings and verification of installed measures.

The following diagram provides an overview of how the Protocols work in relationship to each other and the organizations that are responsible for using the Protocols to conduct evaluation research.

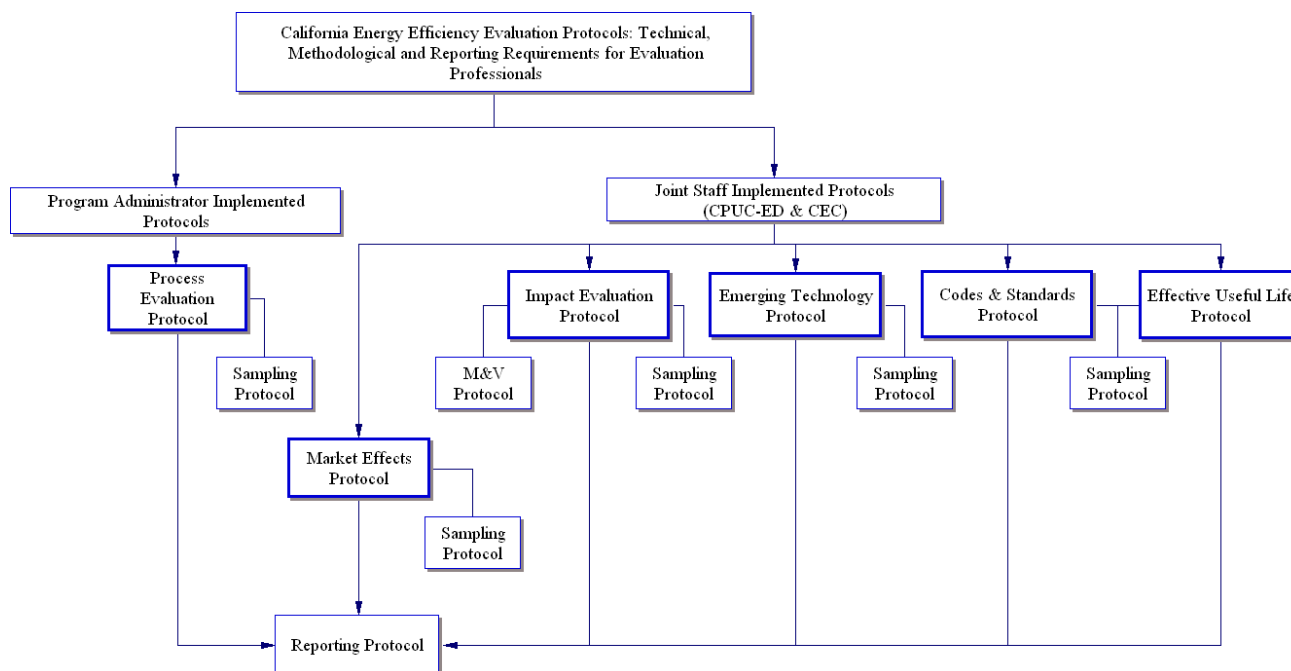


Figure 1. Operational Overview of How the Protocols Relate to Each Other

Note: The Process Evaluation Protocol is a guidance document and is less instructive than the other Protocols that are more prescriptive in design. While the Process Evaluation Protocol does contain required reporting and planning activities, it designates that the key decisions on what, when and how to evaluate are the responsibility of the Administrators.

How the Protocols Meet CPUC Goals

The primary evaluation-related goal of the CPUC is to assess net program-specific energy impacts or the market level impacts of the portfolio of energy efficiency services and to compare these results with the assigned energy savings goals. Similarly, the CPUC must be assured that when an evaluation is conducted it can rely on the findings of that research to accurately reflect the energy benefits available to the citizens of California in exchange for the resources spent. As a result, the following goals are incorporated into the operations of the Protocols:

- To identify the annual energy and peak demand impacts associated with each program offered, for which there are expected savings, over the period of time the program measures are projected to provide net participant energy impacts. This will almost always be for a longer period of time than the program funding cycle;

- To identify the annual energy and peak demand impacts associated with major program delivery mechanisms (e.g., direct install approaches, incentive and rebate approaches, and education, marketing and outreach programs) over the period of time the program measures are projected to provide net participant energy impacts;
- To estimate the annual energy and peak demand impacts associated with each Administrator's portfolio projected over the period of time the program services are expected to provide net energy impacts;
- To compare the evaluation results across programs, types of programs (program groups) and program portfolios to assess their relative performance and cost-effectiveness;
- To identify under-performing program or program components , so they may be improved or withdrawn from the portfolio of services;
- To understand the potential of programs and program services to cost-effectively increase the supply of energy resources for California citizens;
- To understand how programs or program operations can be modified to improve their performance and the overall performance of the portfolios;
- To inform future updates to ex-ante energy and peak demand savings estimates for program planning purposes;
- Provide timely information to improve program design and selection for future program cycles;
- To be able to tailor the evaluation approaches and budgets to meet the need for reliable energy impact and market effects information while minimizing evaluation costs and reducing risks of making poor efficiency supply decisions; and
- To use an objective and transparent evaluation process that assesses the impacts from all types of programs that are expected to provide efficiency resources in California.

The Energy Action Plan, the Energy Efficiency Policy Manual and other related CPUC documents have established aggressive goals for energy efficiency in California. Throughout these guidance documents, it is explicitly recognized that investments in energy savings are uncertain and, hence, carry some risk. The guidance documents emphasize the need for “reliable” savings estimates. Efforts to define “reliable” lead to quantification. To quantify and manage these risks, one must include all relevant and cost-effective sources of information on the performance of the investment and the underlying uncertainty in these data.

To the greatest extent possible, the Joint Staff will seek to allocate evaluation resources to reduce uncertainty in the estimates and evaluations of achieved gross and net savings. The criteria for allocating evaluation resources will be influenced by risk considerations associated with a program's designs and operational characteristics, the expected energy savings, the need to minimize uncertainty in the assessment process and the cost to quantify and manage these risks. The overarching theme in the management of the evaluation effort should follow the IQM risk principle: Identify, Quantify and Manage. This principle is based on the recognition that all estimated savings from energy efficiency and conservation programs (as well as estimated energy and capacity from traditional supply-side resources) include some uncertainty and,

consequently, risk. In the past, planners, evaluators and other staff have often relied on single-point savings calculations (e.g., average kWh savings) that were subsequently discounted, based on professional judgment. Risk was not quantified, therefore it could not be effectively managed. By explicitly identifying factors that induce or affect uncertainty and by taking steps towards quantifying that risk, the Joint Staff can make more informed decisions on how to effectively manage the evaluation efforts and reduce the overall risk associated with the efficiency portfolio.

Use of the Evaluation Results to Document Energy Savings and Demand Impacts

There are several Protocol-guided evaluations that provide net energy impacts that will be used to understand program, portfolio and/or statewide energy savings. These are:

- The direct program impact evaluations that document the energy savings associated with the actions taken through program participation, such as when a rebated motor is installed or when a high-efficiency cooling system is upgraded;
- The indirect program impact evaluations that document the behavioral change, and in some cases the energy savings associated with the behavioral changes made as a result of program activities, such as when training is provided to customers. For example, when a customer installs an energy-efficient technology due to exposure to a training program and without any other program assistance;
- Evaluations conducted according to the Codes and Standards Program Evaluation Protocol that provide the net energy impacts associated with a code or standard change; and
- The market effects evaluations that document the net effects of one or more programs on the operations of a market and applies energy savings estimates to these program-induced market changes.

All of these impacts will be assessed for statewide energy and demand impacts. However, for the purposes of crediting individual programs or Administrator program portfolios with energy impacts, only the first three categories of net energy impacts documented in the evaluations will be counted, and not those from market effects evaluations. The evaluations in the first three categories will derive program-specific net energy impacts and will be used to sum up to the investor-owned utility (IOU) portfolio impacts and used to derive the statewide program impacts.

The Evaluation Identification and Planning Process

The program evaluation planning process shall begin with a high-level assessment of the need to evaluate a program or program component. This assessment will consider, among other factors, the importance of the savings to the portfolio and the uncertainty regarding the ex-ante savings estimates. Based on this assessment, the Joint Staff will decide whether each program or program strategy must comply with the Protocols or whether it will be required to comply only with the CPUC's program reporting requirements.

For those programs that will receive a Protocol-guided evaluation, the next series of issues should be addressed to determine if Protocols that cover multiple types of programs or a program-specific Protocol should be used. These focus on specific types or characteristics of programs. If the program is focused on emerging technologies, then the Emerging Technology Program Evaluation Protocol must guide the evaluation. If it is a Codes and Standards Program, then the evaluation must be guided by the Codes and Standards Program Evaluation Protocol. Other types of program evaluations will be guided by the Protocols designed for a wide variety of resource and non-resource programs.

The next question to address is whether the program or program strategy is expected to obtain *direct* energy or demand savings. Producing savings *directly* means that the link between the program activity and the savings is clear, straightforward and relatively fast. These types of programs are often referred to as resource or resource acquisition programs. An example of such a program is an incentive program, such as a single-family rebate program, that offers incentives to residential customers to install efficient equipment. For each participant who receives an incentive, there is the clear expectation that there will be savings based upon the program's direct results in obtaining equipment installations. Information and education programs are examples of programs that do not provide such direct impacts. For these programs, there is a more tenuous link between the program activities and any eventual savings. That is, a training program may or may not result in any savings and the savings that are achieved are not direct. Savings obtained from providing training services depend upon that program inducing some form of behavior change (such as purchase and installation behavior or participation in a more direct efficiency program). This would be *indirect* savings. If a program is one that provides savings *indirectly*, then its evaluation must be guided by the Indirect Impact Evaluation Protocol that explicitly addresses the need to link program-induced behavioral changes to eventual energy and demand impacts. Some programs may intend to produce energy savings by providing behavior change information or education for which an impact evaluation of energy savings is not needed by the CPUC. These evaluations would follow the Indirect Impact Evaluation Protocol and quantify behaviors changed or actions taken, but not move to the step of allocating energy savings to those efforts. Joint Staff will determine which Evaluation Protocols to apply to which programs as part of their evaluation planning efforts.

If the program is defined as one that *directly* produces energy and demand impacts, it must be determined whether it will be guided by the Impact Evaluation Protocol,³ the M&V Protocol or both. Programs assigned to the M&V Protocol only (not assigned an impact evaluation) will be those for which savings are expected to be relatively small and certain (reliable).

A program with a combination of large and/or uncertain savings must be guided by the Impact Evaluation Protocol. If such programs do not cover any measures that should be specifically evaluated in order to update the Database for Energy Efficiency Resources (DEER) an impact evaluation at the program or program-strategy level (rather than at the technology level) must be planned. However, if the program or program strategy covers measures that should be evaluated

³ The Impact Evaluation Protocol contains the Indirect Impact Protocol and three others related to estimation of direct savings: the Gross Energy Impact, Gross Demand Impact and Participant Net Impact Protocols. The Impact Evaluation Protocol also often "calls for" the M&V Protocol that provides requirements for M&V-related activities within the impact evaluation methods.

in order to update DEER, it must be determined whether there is a *sufficient* number of these measures on which to base a technology-level assessment. If so, evaluators shall develop a measure-level plan to evaluate these technologies, as well as plan an impact evaluation at the program or subprogram level.

If there is an *insufficient* number of a particular measure within a single program, a determination needs to be made whether there is a sufficient number of the measure across the program strategies being addressed within a program group to allow for an evaluation. If so, the evaluator shall develop a measure-level plan to evaluate these technologies. Note that measure-level plans should always be nested within the overall impact evaluation for the program or program strategy. Ultimately, the evaluator must account for all the energy and demand impacts for a given program or program strategy.

Figure 2 illustrates the high-level overview of the program evaluation planning process for programs, program strategies and measures.

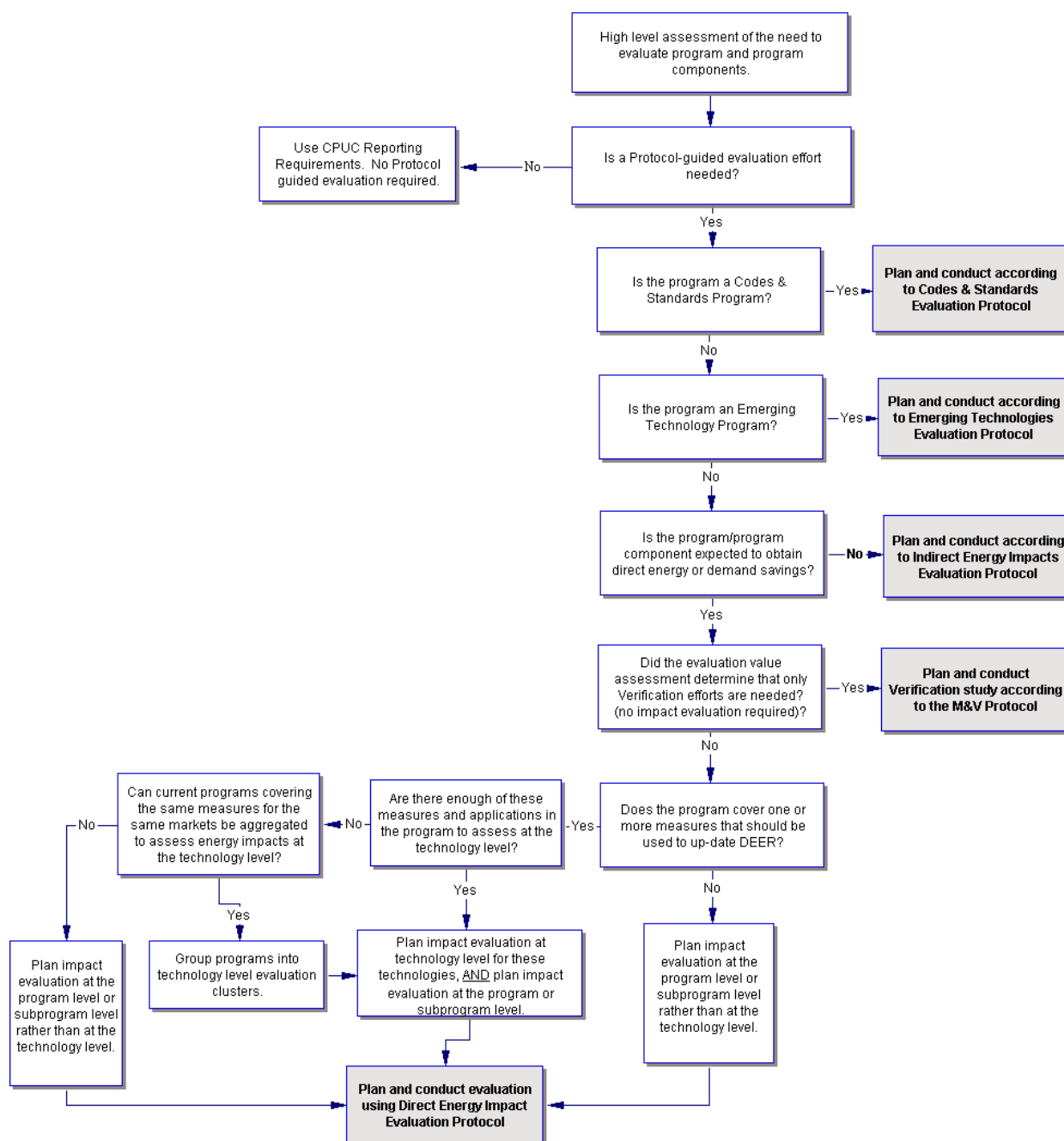
Evaluation Planning Process For Programs, Program Components, And Program Covered Technologies

Figure 2. The Program Evaluation Planning Process for Programs, Program Components and Program-Covered Technologies

The procedure is much less structured for determining when to conduct a market effects study. Figure 3 provides a diagram of the related decision process. In this process, the Joint Staff will examine the mix of programs and strategies within the Administrator portfolios and the markets in which they are operating. Markets will be selected for Market Effects Evaluation when the

Joint Staff finds that such an evaluation would provide valuable information for directing program improvements and/or for better assessing the complete impacts from the portfolio of programs. Markets may be selected for a Market Effects Evaluation due to a preliminary assessment that there are substantial investments in that market across programs where potential market effects (including non-participant spillover) could be measured or need to be tracked and/or assessed. Markets can also be selected for a Market Effects Evaluation when one or more programs operating in that market are best evaluated at the market-level due to their overlapping nature or overlapping goals to change how a market operates (sometimes called market transformation goals).

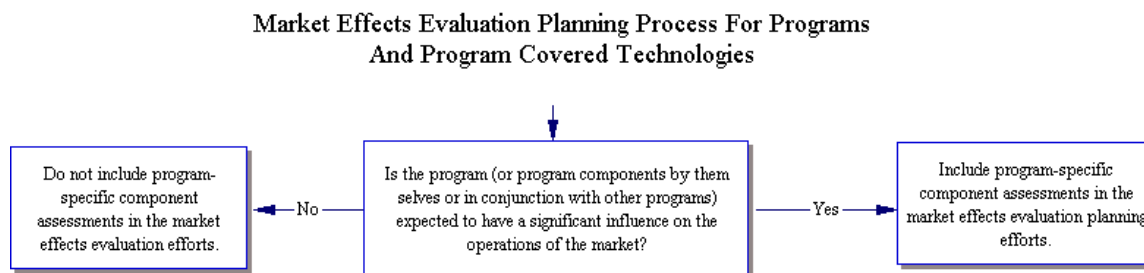


Figure 3. The Market Effects Evaluation Planning Process

Evaluation Rigor and Budgets

The process of setting evaluation priorities and budgets for each type of evaluation effort is as follows:

Impact Evaluations

For impact studies, the Joint Staff will review the Administrator's portfolios and programs and establish evaluation groupings. These groupings will consist of multiple programs having common characteristics that provide evaluation efficiencies in the contracting, supervision and implementation of the evaluation efforts. The groupings will typically include similar types of programs (e.g., residential rebates, commercial rebates, information and education, and marketing and outreach) or markets, so that the evaluation contracts will focus on similar types of programs and program evaluation efforts.

Once the evaluation groups are structured, the Joint Staff will decide which programs (or program components) will receive verification-only analysis, direct impact evaluation or indirect impact evaluation.⁴ Each of these will be assigned minimum rigor level requirements along with a budget based on a number of factors listed in the *Evaluation Framework* including:

- The amount of savings expected from each program in the group;
- Whether the programs are expected to grow or shrink in the future;

⁴ See the Impact Evaluation Protocol herein for further description of these different types of evaluations and the various protocols and rigor levels within them.

- The uncertainty about expected savings and the risk programs pose to achieving portfolio savings goals; and
- How long it has been since the last evaluation and how much the program has changed in the interim.

In setting the level of rigor and the evaluation budgets for the program groups and the individual programs within each group, the Joint Staff will conduct an evaluation needs assessment to assign a level of evaluation rigor to each program or program component. Based on the analysis and criteria listed above, the Joint Staff will establish appropriate evaluation budgets across the program evaluation groups. These budget levels will be used in the development of Request for Proposals (RFPs) to conduct the evaluation efforts. They will also serve to communicate to evaluation contractors how evaluation efforts will be structured.

From this effort, the Joint Staff will provide a high-level evaluation plan that presents the overall evaluation goals and approaches selected for the program groups. The plan will be updated annually as the evaluations proceed, as the need for information changes and as adjustments to the evaluation rigor or approach are identified. The plans will be presented to the public for review and comment each year prior to their implementation in a public workshop to solicit comments and recommendations from interested stakeholders. Once public comments have been obtained, the plan will be finalized and used to support the evaluation bidding and contracting process.

Once an evaluation is launched, the Joint Staff will monitor evaluation efforts and their progress to ensure that evaluation approaches meet or exceed the evaluation rigor assigned, in order to obtain the most reliable evaluation results within the available budgets.

Process Evaluations

For process evaluations, Administrators are responsible for setting evaluation priorities, budgets, evaluation timing and conducting the evaluation effort. These activities are presented to the CPUC-ED, the CEC and the public via an annual portfolio/program evaluation plan and a public workshop. See the Process Evaluation Protocol for additional details.

Market Effects Evaluations

The Joint Staff is responsible for identifying markets for which market effects evaluations will be conducted. These studies will be planned and budgeted individually in accordance with the information and data reliability needs of the Joint Staff.

Codes and Standards and Emerging Technology Program Evaluations

These two program types require evaluations different enough in their goals and objectives, approaches for accomplishing goals and operational characteristics that this document contains Protocols specifically designed for them. While these two types of programs will be evaluated per their respective Protocols, they may also have other types of evaluation efforts applied, such as process or market effects evaluations.

Evaluation Budgets

Each program group evaluation will have a budget cap within which to carry out a variety of evaluation activities. Efforts to maximize reliability will be carried out within the budget constraints and inevitably involve a number of tradeoffs regarding precision and identifying, mitigating and minimizing potential bias. Additional information and guidance on establishing evaluation budgets is provided in the *Evaluation Framework*⁵.

Recommendations for Using the Protocols

The Protocols provide guidance and requirements for planning and conducting California's energy efficiency program evaluations. The Protocols should be used by the Joint Staff and Administrators to structure the evaluation process and associated activities. Joint Staff involved in program evaluation efforts should have an expert understanding of the Protocols. Evaluation staff within Administrator organizations should have the same level of understanding of the Protocols as appropriate to activities in which they have responsibility. All evaluation contractors should be required to have an expert understanding of the Protocols that will directly affect the studies and the methodological approaches they must conduct. It is also recommended that all of these involved parties have a working knowledge of the contents of the *Evaluation Framework* as applicable for the areas in which they work.

When a conflict exists between the *Evaluation Framework* or other reference documents and the Protocols, the Protocols will take precedence unless otherwise approved by the CPUC-ED.

The Detailed Evaluation Work Plan

All program evaluations are required to have a detailed evaluation work plan. In many cases the program evaluation work plans will be clustered within evaluation groupings. However, even within these groupings, there must be a detailed evaluation work plan structured at the program (and in some cases at the program component) level that identifies how the program will be evaluated and the steps to be taken to conduct the evaluation. The evaluation work plan shall include the following components to support an assessment of the adequacy and approach of the evaluation effort:

- Cover page containing the names of the program(s), Administrators and evaluation contractors, date of the evaluation work plan and the program tracking number(s) for program(s) covered in the plan;
- Table of Contents;
- High-level summary overview of the programs and the evaluation efforts;
- Brief description of the program(s) being evaluated including a high level presentation of the program theory. If the program does not have a formal program theory, the evaluation plan should incorporate a brief presentation of the evaluation-assumed program theory so that the Joint Staff may understand the sequence of events leading from program actions and activities to desired results (direct or indirect energy impacts);

⁵ TecMarket Works, 74-79.

- Presentation of the evaluation goals and the detailed researchable issues to be addressed in the evaluation. (These will also be presented and discussed in the evaluation reports;)
- Description of how the evaluation addresses the researchable issues, including a description of the evaluation priorities and the use of assigned rigor levels to address these priorities;
- A discussion of the reliability assessment to be conducted, including a discussion of the expected threats to validity and sources of bias and a short description of the approaches planned to reduce threats, reduce bias and increase the reliability of the findings and minimize bias and uncertainty;
- Task descriptions of the evaluation efforts;
- Description of the analysis activities and approaches to be taken:
 - For energy acquisition and procurement programs, include a description of the approach that will be used to estimate kW, kWh and therm impacts for each year over the EUL of program-covered measures, including a description of the approach to be used to adjust the expected impacts for the persistence of the impacts;
 - For information or education programs, include a discussion of the approach that will be used to estimate the actions or behaviors taken and/or knowledge gained that is expected to lead to energy impacts;
 - For process or operational assessments, include a description of the approach used to identify changes that can be expected to improve the cost-effectiveness of or participant satisfaction with the program;
- Description of the M&V efforts (impact evaluations only) including:
 - Reference to International Performance Measurement and Verification Protocol (IPMVP) option⁶, if used;
 - Detailed description of the option-specific approach; and
 - Description of any deviations from the IPMVP option, if any;
- Description of the sampling rationale, methods and needed sample sizes.
- Discussion of the specific Performance Basis Metrics that will be reported in the draft and final evaluation plan;
- A definition of the terms “participant” and “non-participant” as it applies to the evaluation being conducted;
- Detailed description of the information that will be needed from the IOUs or from the program-reporting database maintained at the CPUC-ED in order to conduct the evaluation and an estimate of the date that the information will be needed. This same information will be included in evaluation-related data requests;

⁶ More information on the IPMVP can be found in the *Evaluation Framework* (148-149), or at the IPMVP Web site at www.ipmvp.org.

- Evaluation activities timeline for the program cycle, including identification of deliverables and deliverable due dates. This should also include early, mid-stream and late cycle feedback deliverables and deliverable dates. (These dates must be coordinated with the information needs of the Joint Staff and their program-portfolio assessment needs schedule;)
- Total program budget, total evaluation budget and a task-level evaluation budget for the study; and
- Contact information for the lead Administrator, lead program manager and evaluation manager, including addresses, telephone numbers, fax numbers and e-mail addresses.

The evaluation work plan should be written in a style and with enough detail that it can be clearly understood by Administrators, policy makers and evaluation professionals, and replicated by other evaluation contractors.

Confidentiality Issues

Confidentiality is an essential part of the evaluation process and is included in this section to set a baseline for how information will be treated within the evaluation efforts. The following aspects of confidentiality are incorporated into all evaluations conducted under the guidance of the Protocols.

1. All evaluation contractors will be required to sign confidentiality agreements in order to conduct evaluations funded through the Protocols. These agreements will be incorporated into all evaluation contracts. For impact, market effects, codes and standards, emerging technology and M&V studies, the agreements will be incorporated into contracts awarded by the CPUC or the CEC as appropriate. For process evaluations, the individual Administrators issuing the process evaluation contracts are responsible for incorporating confidentiality agreements. However, evaluation information, including customer-specific information, can be shared across evaluation contractors within the same evaluation team and across teams. However, this data is to be protected from exposure beyond the evaluation teams and all contractors must sign confidentiality agreements prior to the receipt of customer-specific information.
2. All customer-specific information will be treated as confidential and safeguarded from public disclosure. Evaluation contractors are granted access to participant and customer specific information maintained by the Administrators as needed to conduct the evaluation efforts, however, no evaluation contractor will allow participant or customer specific information to be released to individuals or organizations beyond their research team, unless specifically permitted in writing by each customer for which information is to be released. All memoranda, letters, e-mails, reports and databases that are developed or used in the evaluation efforts that contain participant-specific or customer-specific information, whether an individual, a firm or business or an organization, are covered by this confidentiality requirement.

Contacting the Customer

A critical component to the success of any evaluation effort is the maintenance of a supportive relationship between the customer and the many different types of organizations that influence the evaluation effort. IOU representatives, CPUC-ED, CEC, evaluation contractors and others involved in the evaluation efforts need to be diligent in making sure that customers and participants are not over-contacted in support of them. Whenever possible, customer contact initiatives should be coordinated to avoid over-contact. Customer requests to be excluded from evaluation efforts should be respected. Customer complaints associated with evaluation efforts should be reported to the CPUC-ED and the associated Administrator within 48 hours of receipt.

Before customers are contacted by evaluation contractors, their representatives or subcontractors, the prime evaluation contractor will notify the Administrators of the need to do so and work to agree on an approach and timeline that may change from study to study. All final customer contact approaches and contact Protocols should specify customers to be contacted (as an attachment), reasons for the contact, information to be collected, the method of contact and the associated timeline.

Administrators will inform the appropriate individuals within their organizations of any related customer contact.

Impact Evaluation Protocol

Introduction

The Impact Evaluation Protocol is applicable for all programs or program components designated by the Joint Staff for a direct or indirect⁷ impact evaluation, especially for those programs claiming energy or demand savings and for those programs that are expected to influence energy-related behaviors and can be linked to energy and/or demand savings. This Protocol is designed to reliably estimate program impacts. Information, education and advertising efforts determined by the Joint Staff to have an indirect impact evaluation are expected, at a minimum, to measure the program-induced behavioral changes, often leading to energy and demand savings estimates.⁸

The Impact Evaluation Protocol is established to ensure that all evaluations of program-specific energy and demand savings, and program-specific impacts are conducted using evaluation methods deemed acceptable based on the assigned level of rigor for that evaluation. The Protocol's list of allowable methods is one component that helps ensure greater reliability in the energy and demand savings estimates from California's energy efficiency efforts. The Joint Staff can assign different levels of rigor to each program, thus allowing the flexibility to allocate evaluation resources according to the needs of the Portfolio given uncertainty in the expected savings, the size of expected savings, the program budget and other criteria. The Joint Staff will instruct evaluation contractors to use specific rigor levels based on its application of the Protocol's decision criteria, and mix of evaluation choices and resource allocations.

Rigor is defined as the level of expected reliability. The higher the level of rigor, the more confident we are that the results of the evaluation are both accurate *and* precise, i.e., reliable. That is, reliability and rigor are synonymous. Reliability is discussed in the Sampling and Uncertainty Protocol and in the *Evaluation Framework* where it is noted that sampling precision does not equate to accuracy. Both are important components in reliability, as used by the CPUC. Each program will be assigned a specific evaluation rigor level for its primary evaluation objectives to guarantee that a minimum standard is met.

"Impact evaluation" refers here to all program-specific evaluations designed to measure program impacts. Impact evaluations attempt to estimate net changes in electricity usage, electricity demand, usage of therms and/or behavioral impacts that are expected to produce changes in energy use. Evaluations conducted according to the Impact Evaluation Protocol are expected to obtain energy or demand savings estimates wherever possible. Impact evaluations of programs or program components designed to directly achieve energy and demand savings should follow the Direct Impact Evaluation Protocol to measure these savings.

⁷ The term "indirect impact evaluation" refers to those program-specific evaluations designed to measure the specific program goals that create an impact that is expected to eventually lead to energy and/or demand savings but where these savings cannot be directly estimated.

⁸ This is the minimum expectation. The evaluation research design, however, could surpass this through an experimental or quasi-experimental design that estimates energy and demand savings.

The Indirect Impact Evaluation Protocol should be used for evaluations of those programs or program components primarily designed to obtain behavior changes that will eventually lead to energy and demand savings but do not directly do so within the program. The Indirect Impact Evaluation Protocol is intended for those programs where the primary uncertainty lies in the program's ability to obtain the behavior change(s) targeted by the program. Indirect impact evaluations will, therefore, be linked wherever possible to previously measured energy or demand savings estimates that would yield savings estimates with the same rigor required by the Basic rigor level for impact evaluations described below. This link to reliable stipulated or engineering calculated energy and demand savings estimates is not always possible for behavioral program efforts assigned to receive an impact evaluation. In these cases, an indirect impact evaluation shall be conducted at the Basic rigor level to measure the program-induced impacts, as described below.

This Protocol often refers to a "program or program component." A program component, as defined by this Protocol, is any identifiable portion of a program. This could be a measure, a delivery mechanism, a set of delivery mechanisms or measures, or a set of delivery mechanisms or measures that follow a chain from an activity depicted in a program logic model. The Joint Staff may desire a direct or indirect impact evaluation of a program as well as a separate analysis for the impact evaluation for one of its program components. This might occur, for example, when more detailed evaluation information is needed for a measure for future program planning or to support an update of DEER, a new measure is being piloted or expanded in its use, or a new delivery mechanism has been added.

Impact evaluations are limited to addressing the direct impacts of the program on participants and estimating participant spillover impacts.⁹ These studies do not include documenting program influences on the operations of a market or the program's impacts on non-participants. Program-induced changes on the way a market operates or on non-participants are addressed in the Market Effects Evaluation Protocol.

The Impact Evaluation Protocol describes the metrics to be produced from an impact evaluation. This includes the target parameters that must be used as part of developing the evaluation design in order to produce these metrics. This Protocol also presents an overview of how the Impact Evaluation Protocol is integrated with the M&V, Sampling, Market Effects and Reporting Protocols for the implementation of a direct impact or indirect impact evaluation and within the overall system to produce reliable portfolio level evaluated savings estimates. This systematic Protocol-linked process is designed to be part of a proposal selection and evaluation plan review process, which is followed by ongoing management of the evaluation and evaluation reporting.

It is expected that evaluation contractors will respond to requests for proposals (RFPs) for impact evaluations with proposals that meet the standards contained in the Protocols. It is expected that generally accepted statistical methods as published in textbooks used at accredited universities or articles in peer-reviewed journals will be used for parameter estimation from sample data in

⁹ For a thorough evaluation, impact evaluations should estimate direct program savings and participant spillover savings. These estimates need to be distinct estimates and not a combined estimate across the two whenever possible. Current CPUC policy, as the Protocols are being developed, states that only program savings and not participant spillover will be counted towards program and administrator goals and performance.

regression-based approaches and for moving from sample estimation to program or population-based estimates. Engineering methods are expected to meet the requirements in the M&V Protocol and follow generally accepted practices as published in engineering textbooks used at accredited universities or articles in peer-reviewed journals along with generally accepted statistical methods (as described above). Evaluation contractors may propose optional methods in addition to Protocol-compliant methods, if the optional methods provide at least as much rigor and accuracy as the Protocol-covered approach.

Audience and Responsible Actors

The audience and responsible actors for this Protocol include the following:

- **Joint Staff Evaluation Planners** will use the Protocol to determine when a direct impact or indirect impact evaluation is appropriate and to assign the level of rigor expected for the study, as input into the evaluation RFPs for impact evaluation contractors, and as background and criteria for use in reviewing impact evaluation plans, managing the impact evaluations and reviewing impact evaluation reports and results;
- **The Evaluation Project Team** will use the Protocol to ensure that their detailed direct impact or indirect impact evaluation plan(s) address(es) key requirements for each program or program component based upon the level(s) of rigor designated by the Joint Staff. They will also use the Protocol to double-check that the Protocol requirements have been met as they conduct, complete and report the impact evaluations;
- **Administrators** will use the Protocol to understand how the impact evaluation will be conducted on their programs and to understand the evaluation data needs to support the impact evaluation. In addition, the Protocol provides background for the Administrator's use to determine when to intervene in the program design and implementation efforts to achieve continued and/or greater efficiency gains;
- **Program Implementers** will use the Protocol to understand the impact evaluation that will be conducted of their programs and program components. Often, they will be required to provide data to support the impact evaluation. The Protocol will also provide background for their use to understand when to intervene to achieve continued and/or greater efficiency gains; and
- **ISO / System planners** will use savings and uncertainty estimates for load forecasting and system planning.

Overview of the Protocol

Protocol Types

The overall Impact Evaluation Protocol contains one subset of 3 Protocols for estimating direct energy and demand impacts and one for estimating indirect impacts.

Direct Impact Evaluation Protocols:

- The Gross Energy Impact Protocol has two levels of rigor (Basic and Enhanced) for developing gross energy estimates;

- The Gross Demand Impact Protocol has two levels of rigor (Basic and Enhanced) for developing gross demand estimates; and
- The Participant Net Impact Protocol has three levels of rigor for developing net impact estimates (Basic, Standard and Enhanced).

The Indirect Impact Evaluation Protocol has three levels of rigor (Basic, Standard and Enhanced). The Basic Rigor level is reserved for those programs or program components that cannot be linked to energy savings but where net behavior changes need to be estimated to measure program impacts. This Protocol includes the requirement that the measured impacts are net impacts (i.e., program-induced).

Rigor

The general rules for how often evaluations need to be conducted are determined by the Joint Staff. The Joint Staff will decide, for each relevant program, if and when the program will receive an impact evaluation. The Joint Staff may choose not to have an impact evaluation conducted for a particular program or program component. When the Joint Staff decides a program will receive an impact evaluation, it also selects whether a direct impact evaluation or indirect impact evaluation is most appropriate and the level of evaluation rigor required. The Impact Evaluation Protocol then establishes the methods appropriate for the given type of impact evaluation and assigned level of evaluation rigor. In this way, the Protocols establish a minimum level of evaluation rigor in order to ensure that the savings estimates produced are at the level of reliability needed to support the overall reliability of the savings in the Administrator's Portfolio and the statewide Portfolio.¹⁰

Each level of rigor provides a class of allowable methods in order to offer flexibility for the potential evaluation contractors to assess and propose the most accurate and cost-effective methods that meet the Joint Staff's needs. The principle is to provide minimum specifications for a set of options at each rigor level and yet encourage evaluation contractors to use both the art and science of evaluation to develop affordable and quality evaluations that produce reliable savings estimates.

The Joint Staff may assign one rigor level for a program and a different level of rigor for one or more of its program components. When this happens, the evaluation must meet the level of rigor for that program component as assigned (to include meeting the Sampling and Uncertainty requirements) as well as the rigor level for the program as a whole.

The various Protocols and associated rigor levels required for direct impact and indirect impact evaluations are illustrated in Figure 4.

¹⁰ Savings for programs with expected savings could be included in the Portfolio savings estimates based upon an accounting effort that multiplies the number, or verified number, of installations times the latest evaluated savings estimates or deemed savings, as determined more appropriate by Joint Staff. The verified number of installations is the number of program installations based upon the reported number combined with the results of any verification activities required by Joint Staff. (See the M&V Protocol for the description of the Verification Protocol.)

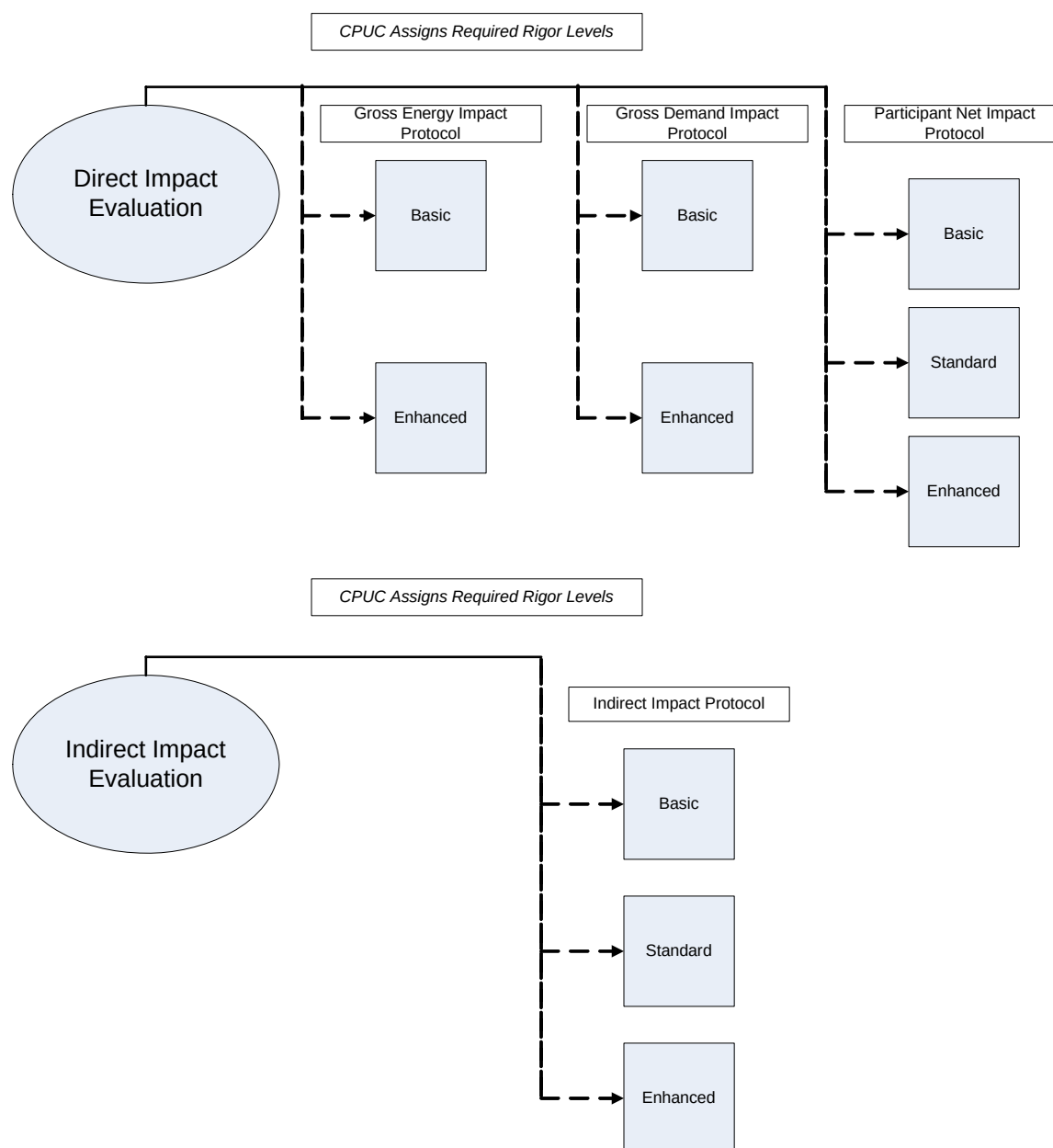


Figure 4. Required Protocols for Direct Impact and Indirect Impact Evaluations

Key Metrics, Inputs and Outputs

Impact evaluations will draw upon data from program databases, program descriptions, DEER databases, work papers developed during program planning, utility demand metering and consumption data for participants and non-participants, utility-, state government- or local government-collected weather data, on-site measurement, monitoring and observational data, survey and interview data collection, and other prior study data and reports. These will be used

with the Impact Evaluation Protocol-allowable methods to produce program, program strategy and program component (measure-level, as requested) impact evaluations. These must be conducted using the Joint Staff-approved evaluation plans.

The Impact Evaluation Protocol will guide the estimation of evaluation-adjusted gross and net savings for energy (kWh) and demand (kW) for electricity-using equipment (and behaviors related to electricity-using equipment) and net therm savings for gas-using equipment (and behaviors related to gas-using equipment). The kWh, kW and therm impacts are required to be reported separately for the first year and for each year thereafter over the period of time in which net program-induced savings are expected. The programs' expected savings from program plans, reported savings and the evaluation's estimate of savings will be reported in these annual savings tables. The Reporting Protocol, which all direct impact and indirect impact evaluation reporting must follow, provides further description and table examples.

Because impact evaluations must follow the Sampling and Uncertainty Protocol, evaluators must also assess, minimize and mitigate potential bias and present the achieved level of precision (including relative precision, error bounds, coefficient of variations, standard deviations and error ratios) for interpreting information, summarizing savings and its precision across programs, and providing the information necessary for future evaluation planning. Where precision is calculated from the chaining or pooling of evaluation study efforts, the above precision information should be provided for each study effort as well as the combined result.

When requested by the Joint Staff, impact evaluations must produce the required metrics by delivery mechanism (e.g., rebates and direct install). Where delivery mechanisms differ within a program, this Protocol requires that the impact evaluation be designed, conducted and reported to provide the energy and demand metrics (along with the precision information) for each delivery mechanism, when the Joint Staff identifies delivery method-associated impacts as an evaluation goal.

Evaluations conducted according to the Gross Impact Protocol, Gross Demand Protocol and Participant Net Impact Protocol will produce gross and net kWh, kW and therm impacts. The evaluation analysis results must be used with program database, verification, standard weather information, and other participant and non-participant data, as necessary, to produce program energy and demand savings estimates. Measure effective useful life (EUL) from DEER or as otherwise approved by the Joint Staff will be used to create the required energy and demand impacts for first year and for each year thereafter over the period of time in which savings are expected based upon measure EUL. Any evaluation findings that might call into question the EULs being used must be presented to the Joint Staff when discovered and discussed in the evaluation report. Further description and examples of the required tables are provided in the Reporting Protocol.

All direct impact and indirect impact evaluations are expected to assess and discuss the differences between the ex-ante estimates and the evaluation produced ex-post estimates. To the extent that the data gathered and evaluation analyses conducted can explain the causes for these differences, this must be presented and discussed. Cases in which explaining these differences due to lack of data or problems of interpretation should be noted in the evaluation report.

Energy and Demand Impact Protocols

These are minimum standard Protocols. All methods in a higher class of rigor are allowable as they exceed the minimum criteria. For example, if the program has a Joint Staff assigned rigor of Basic or Standard and the method proposed by the evaluation contractor is an option under a rigor of Enhanced (but is not listed under Basic or Standard), this method is acceptable for meeting the Protocol.

Gross Energy Impact Protocol

The Gross Energy Impact Protocol is summarized in Table 1. Further description, additional requirements, clarification and examples of this Protocol are presented after the table. The methods used and the way in which they are used and reported must meet all the requirements discussed within this section (not just those within the summary table or those within the text) to provide unbiased reliable estimates of program level gross energy impacts in order to comply with the Gross Energy Impact Protocol. The Protocols sometimes reference the *Evaluation Framework* or other documents which provide examples of applicable methods. The requirements, however, are always those stated in the Protocols, which take precedence over all other protocols and evaluation guidance documents in all circumstances, unless otherwise approved or required by the CPUC.

All M&V referred to in the Impact Evaluation Protocol must be planned, conducted and reported according to the M&V Protocol. M&V may be conducted at a higher level of rigor, with more inputs measured or metered, or with greater precision than the minimum shown within the Impact Evaluation Protocol, but not with a lower level of rigor. The M&V Protocol can also be required by the Joint Staff or used by evaluators to enhance other evaluation efforts. For example, an evaluator proposing a Statistically Adjusted Engineering (SAE) regression model may use (or the Joint Staff may require the use of) the M&V Protocol to conduct field measurements on the sample of participants to be included in the SAE model to improve the engineering estimates. This may involve conducting measurement/metering and utilizing IPMVP Option A.

The overall goal of the Direct Impact Evaluation Protocol (which includes the Gross Energy Impact Protocol) is to obtain unbiased reliable estimates of program-level net energy and demand savings over the life of the expected net impact.

Table 1. Required Protocols for Gross Energy Evaluation

Rigor Level	Minimum Allowable Methods for Gross Energy Evaluation
Basic	1. Simple Engineering Model (SEM) with M&V equal to IPMVP Option A and meeting all requirements in the M&V Protocol for this method. Sampling according to the Sampling and Uncertainty Protocol. 2. Normalized Annual Consumption (NAC) using pre- and post-program participation consumption from utility bills from the appropriate meters related to the measures undertaken, normalized for weather, using identified weather data to normalize for heating and/or cooling as is appropriate to measures included. Twelve (12) months pre-retrofit and twelve (12) months post-retrofit consumption data is required. Sampling must be according to the Sampling and Uncertainty Protocol.
Enhanced	1. A fully specified regression analysis of consumption information from utility bills with inclusion/adjustment for changes and background variables over the time period of analysis that could potentially be correlated with the gross energy savings being measured. Twelve (12) months post-retrofit consumption data are required. Twelve (12) months pre-retrofit consumption data are required, unless program design does not allow pre-retrofit billing data, such as in new construction. In these cases, well-matched control groups and post-retrofit consumption analysis is allowable.¹¹ Sampling must be according to the Sampling and Uncertainty Protocol utilizing power analysis as an input to determining required sample size(s). 2. Building energy simulation models that are calibrated as described in IPMVP Option D requirements in the M&V Protocols. If appropriate, may alternatively use a process-engineering model (e.g., AirMaster+) with calibration as described in the M&V Protocols. Sampling according to the Sampling and Uncertainty Protocol. 3. Retrofit Isolation engineering models as described in IPMVP Option B requirements in the M&V Protocols. Sampling according to the Sampling and Uncertainty Protocol. 4. Experimental design established within the program implementation process, designed to obtain reliable net energy savings based upon differences between energy consumption between treatment and non-treatment groups from consumption data.¹² Sampling must be according to the Sampling and Uncertainty Protocol.

¹¹ Post-retrofit only billing collapses the analysis from cross-sectional time-series to cross-sectional. Given this, even more care and examination is expected with regard to controlling for cross-sectional issues that could potentially bias the savings estimate.

¹² The overall goal of the Direct Impact Protocols is to obtain reliable net energy and demand savings estimates. If the methodology directly estimates net savings at the same or better rigor than the required level of rigor, then a gross savings and participant net impact analysis is not required to be shown separately.

Basic Rigor

There are two classes of evaluation methods that set the minimum allowable methods for the Gross Energy Impact Protocol Basic rigor level.

Simple Engineering Model (SEM)

The first class of allowable methods is the simple engineering model (SEM). An SEM is equivalent to IPMVP Option A and must be conducted as described in the M&V Protocol. This method is described and a few references are presented in the *Evaluation Framework*¹³. These types of models can be straightforward algorithms for calculating energy impacts for non-weather dependent measures such as energy-efficient lighting, appliances, motors and cooking equipment. Exceptions to this requirement are programs offering comprehensive measure packages with significant measure interactions, to include commissioning and retro-commissioning programs, and new construction programs. Evaluations of these programs conducted using engineering methods must use building energy simulation modeling under IPMVP Option D as described in the Enhanced rigor level of the Impact Protocol.

Sampling for the M&V used in the SEM must be conducted as prescribed in the Sampling and Uncertainty Protocol which includes developing the sample to target a minimum of 30 percent precision at a 90 percent confidence level. Knowledge of the components of the SEM and the propagation of error method must be used to determine what needs to be measured in the SEM to meet this requirement. (See the M&V Protocol for more detail on the related requirements.) In both the evaluation plan and the evaluation report, the inputs selected and the methods selected for the measurement/monitoring must be justified in terms of why they are the factors that provide the most likely unbiased and reliable gross energy impact estimates for the evaluation study being conducted.

Normalized Annual Consumption (NAC)

The second class of allowable methods is normalized annual consumption (NAC) analysis. This is a regression-based method that analyzes monthly kWh or therm consumption data provided by utilities. This method and a few references are presented in the *Evaluation Framework*¹⁴. The NAC analysis can be conducted using statistical software, such as the Princeton Scorekeeping Method (PRISM), and other statistically based approaches using SAS or SPSS. The NAC method, often using PRISM, has been most often used to estimate energy impacts produced by whole house retrofit programs.

To comply with this Protocol, NAC must normalize consumption for weather effects using a generally accepted set of weather data (from utility weather monitoring stations, National Oceanic & Atmospheric Administration (NOAA) weather monitoring stations or others as used by California energy forecasting and supply analysts). Weather data must be used to normalize for heating and/or cooling as appropriate to the measures included. Final savings estimates must also use weather data to report both actual savings from the weather data used in the analysis and expected annual savings fitted to the CEC climate thermal zone (CTZ) long-term average weather data.

¹³ TecMarket Works, 123-129.

¹⁴ Ibid, 105-106.

A minimum of twelve months pre-retrofit and twelve months post-retrofit consumption data is required. However, there might be a number of participants who are excluded from the analysis because they do not have the required minimum of twelve months of pre- and post-consumption data. For example, some populations, because they are more mobile (e.g., rental populations and particularly low-income households), will be less likely to have the required amount of pre- and post-consumption data. An examination should be made on whether the inclusion or exclusion of such participants could potentially bias the results.

Often, a census approach is undertaken for NAC. Where sampling is used, it must follow the Sampling and Uncertainty Protocol.

Enhanced Rigor

There are four classes of allowable methods to meet the minimum requirements for the Gross Energy Impact Protocol Enhanced rigor level. One of these is regression analysis of consumption with specific modeling requirements, two are different engineering-based methods with specific M&V and model calibration requirements, and the fourth is experimental design established within the program implementation process to specifically obtain unbiased reliable estimates of net energy and demand savings.

Regression Analysis

The first class of allowable methods is regression analysis of consumption data provided by utilities that statistically adjusts for key variables that change over time and are potentially correlated with gross or net energy savings. As a way of capturing the influence of weather, evaluators may incorporate weather-normalized consumption as the dependent variable or include heating- and cooling-degree days directly in the model. Other variables that change over time that are often correlated with gross and net energy savings include, among others, the state of the economy (recession, recovery, economic growth),¹⁵ fuel prices, occupancy changes, behavior changes (set-point changes, schedules, usage frequency), changes in operation and changes in schedule. The evaluator is free to select the most appropriate additional variables to include.

The modeler is also free to select the functional form of the model (a variety of linear and non-linear forms) as well as the type of model. A wide variety of model types may be used, including Statistically Adjusted Engineering (SAE) models, Analysis of Covariance (ANCOVA) or fixed-effects models and other regression models. The *Evaluation Framework* presents the SAE model and a few references,¹⁶ and ANCOVA with a few references.¹⁷ These types of impact evaluations have been conducted for residential whole-house, heating and cooling retrofit, refrigerator and water heating replacement, and small and large commercial programs. The Enhanced Gross Impact regression option is not limited to these two types of models. Finally, the testing of alternative specifications is encouraged.

¹⁵ See the discussion on page 118 in the *Evaluation Framework* and the article cited in its footnote 82 for more information and an example.

¹⁶ TecMarket Works, 108-109.

¹⁷ Ibid, 109-111.

Power analysis,¹⁸ results from prior studies on similar programs, and professional judgment are to be used to determine the required sample size. Sampling and analysis and mitigation for uncertainty must be planned and conducted according to the Sampling and Uncertainty Protocol.

The primary consideration in the use of regression analysis to meet the Enhanced Gross Energy Impact Protocol is that the analysis must be designed to obtain reliable energy savings. In order for regression to begin to meet the unbiased element in this requirement, the regression analysis must incorporate and control for background and change variables that might otherwise bias the measurement of the energy savings. There are several ways in which this can be accomplished. One common method is to include participant and non-participant analyses. If this method is selected, particular care and justification must be made for the non-participant group selected and its appropriateness for the program and participant population being analyzed. Secondly, research design and analysis needs to consider whether the analysis is providing gross impact, net impact or something in between that must then be adjusted or analyzed in a second step to produce, at a minimum, reliable unbiased net of free-ridership savings estimates.

Alternatively, surveys of participants and the creation of change variables can be created and incorporated into the regression analysis. Another example would be to create or obtain participant or non-participant change variables from secondary or other aggregate or individual studies of similar/matched populations for inclusion within the regression analysis. The specific method and research design to accomplish this requirement is not specified, but the evaluation plan, analysis and evaluation report must present, justify, discuss and analyze the method and data utilized to accomplish this requirement.

A minimum of twelve months pre-retrofit and twelve months post-retrofit consumption data is required. However, there might be a number of participants who are excluded from the analysis because they do not have the required minimum of 12 months of pre- and post-consumption data. For example, some populations, because they are more mobile (e.g., rental populations and particularly low-income households), will be less likely to have the required amount of pre- and post-consumption data. An examination should be made on whether the inclusion or exclusion of such participants could potentially bias the results.

Twelve months pre-retrofit billing data are required unless the program design does not allow pre-retrofit billing data, such as in new construction. In these cases, well-matched control groups and post-retrofit billing analysis is allowable. Post-retrofit only billing collapses the analysis from cross-sectional time-series to cross-sectional. Given this, even more care and examination is expected with regard to controlling for cross-sectional issues that could potentially bias the savings estimate.

Final savings estimates must report both actual savings from the weather data used in the analysis and expected annual savings fitted to the CEC CTZ long-term average weather data.

¹⁸ Power analysis is a statistical technique to determine sample size requirements to ensure statistical significance can be found. There are several software packages and calculation Web sites that conduct the power analysis calculation. See the Sampling and Uncertainty Protocol for more discussion and reference. Power analysis is only being required in the Protocol for determining required sample sizes. Appendix D provides further detail on using power analysis for developing sample size requirements.

The regression-based methods must use power analysis to plan their sample size (unless census samples are being used). Regression-based methods must also meet the requirements of the Sampling and Uncertainty Protocol. Many of the requirements in the Sampling and Uncertainty Protocol require specific actions and documentation regarding data cleaning, model specification, testing and reporting for regression-based methods.

Engineering Models

The second class of allowable methods is building energy simulation programs calibrated as described in the Option D requirements in the M&V Protocols. This method is described and a few references are presented in the *Evaluation Framework*.¹⁹

The engineering models that meet the Option D requirements are generally building energy simulation models, as described in the *Evaluation Framework*.²⁰ This can be applicable to many types of programs that influence commercial, institutional, residential and other buildings where the measures impact the heating, ventilation or air conditioning (HVAC) end-use. This method is often used for new construction programs and building, heating/cooling or shell measure retrofits in commercial and residential programs.

In addition, industrial efforts can include changes in process operations and the appropriate type of model could be a process-engineering model. These are specialized engineering models and software that conduct engineering analysis for industry-specific industrial processes. Where these types of models are more appropriate, the Gross Energy Impact Protocol allows the use of a process engineering model with calibration as described in the M&V Protocols to meet the Enhanced rigor level.

Sampling must be conducted according to the Sampling and Uncertainty Protocol.

Retrofit Isolation Measurements

The third class of allowable methods is the retrofit isolation measurements as described in Option B requirements in the M&V Protocols. This method is used in cases where full field measurement of all parameters for the energy use for the system where the efficiency measure was applied are feasible and can provide the most reliable results in a cost-efficient evaluation. An overview of this method is provided in the *Evaluation Framework*.²¹ Applying a variable frequency drive to a constant speed pump in a variable flow pumping application would be a typical example of when this method would likely be used.

Sampling must be conducted according to the Sampling and Uncertainty Protocol.

Experimental Design

The fourth class of allowable methods is experimental design. Experimental design with energy and demand measurement (either consumption data comparison or engineering-based with M&V) comparisons between the treatment and non-treatment groups meets the Enhanced Gross

¹⁹ TecMarket Works, 129-133 and 176-181.

²⁰ Ibid, 176-181.

²¹ Ibid, 166-169.

Energy Impact Protocol rigor level. Experimental design will normally measure net energy and demand impacts and meet the criteria for equal or better rigor for the overall net savings and demand estimates such that the Gross Impact Protocol and the Participant Net Impact Protocol requirements are met. Currently, experimental design has not been widely used within efficiency evaluation. See the *Evaluation Framework*²² for a description and some examples of potential experimental designs within energy efficiency efforts. Sampling conducted as part of the experimental design must be conducted according to the Sampling and Uncertainty Protocol.

All Gross Energy Impact Methods

All impact evaluations should employ a research design that has properly identified participants made available from the program database(s). The regression methods of pre- and post-consumption and the calibrated engineering model equivalent to Option D could yield results not restricted to the program being evaluated if participation in multiple programs occurs around the same time period or overlaps in influence. This could contribute to double counting at the portfolio level. To avoid this possibility, all Administrators are required to provide data on participation in other programs for all program participants, including when participation occurred. Evaluators are required to ensure that their methodologies and analysis account for any overlap in program participation and measures that could potentially bias the program evaluation results.

All impact evaluations must meet the requirements of the Sampling and Uncertainty Protocol. Regression analysis of consumption data requires addressing outliers, missing data, weather adjustment, selection bias, background variables, data screens, heterogeneity of customers, autocorrelation, truncation, error in measuring variables, model specification and omitted variable error, heteroscedasticity, collinearity and influential data points. Engineering analysis and M&V-based methods are required to address sources of uncertainty in parameters, construction of baseline, guarding against measurement error, site selection and non-response bias, engineering model bias, modeler bias, deemed parameter bias, meter bias, sensor placement bias and non-random selection of equipment or circuits to monitor.

Each item in these lists above must be addressed as they all have the potential to bias the savings estimates. Bias is the greatest threat to the reliability of savings estimates. The primary difference between the Basic and Enhanced rigor levels is that the minimum allowable methods in the Enhanced rigor level directly address or control for the more likely sources of potential bias in that class of methods (e.g., regression-based versus engineering-based). This means that the minimum allowable methods in the Enhanced rigor level are expected to provide more reliable savings estimates than the minimum allowable methods in the Basic rigor level.

All impact evaluations must meet the rigor level assigned. If rigor is assigned for a measure or program component, the rigor level must be met for analysis of that measure or program component. If measure-level analyses are conducted and no rigor level has been assigned for these measures, they may be conducted at either the Basic or Enhanced rigor level as long as the impact evaluation of the program as a whole is designed to achieve its overall target precision

²² Ibid, 104-105.

level and addresses all of the potential bias issues listed above and described in the Sampling and Uncertainty Protocol.

Experience in energy efficiency program evaluation has shown that there are cases where some methods are more likely to yield defensible results than others for certain sectors or program designs. The Impact Evaluation Protocol does not restrict the methods used to those that have been successfully used previously. However, the Joint Staff will consider this factor in both contractor selection and in the review and approval process of the evaluation plan. Methods proposed that do not have a successful track record must have thorough documentation on how the methods, techniques or data that will be used can be expected to produce reliable savings estimates and how the key personnel conducting this effort are qualified to do so. For example, experience to date in energy efficiency impact program evaluation has generally shown the following:

- NAC methods are most applicable to residential and small commercial efforts where the expected energy savings are at least 10 percent of pre-installation usage;
- NAC methods are not well suited to handle significant issues with heteroscedasticity, truncation, self-selection or changes in background issues (e.g., significant change in economic conditions-large recession, recovery or economic growth);
- SEM methods are not well suited for whole building measures with interactive effects or commissioning/retro-commissioning efforts;
- The heterogeneity and multitude of background variable issues for industrial customers and unique commercial (e.g., ski resorts and amusement parks/facilities) or institutional (e.g., water/wastewater and prisons) customers make the use of any regression-based consumption analysis difficult and potentially less reliable than engineering-based methods;
- Regression-based consumption analyses are less likely to be able to obtain definitive energy savings estimates where the expected energy savings are not at least 10 percent of pre-installation usage; and
- Regression-based consumption analysis is quite difficult for new construction programs due to the lack of pre-retrofit consumption data and the consequential greater burden for controlling for cross-sectional issues for comparing participants and non-participants (and self-selection bias, particularly if the non-participants are any form of rejecters of program participation). New construction program impact evaluations are generally conducted using engineering models (such as those described in IPMVP Option D).

Gross Demand Impact Protocol

The Gross Demand Impact Protocols are summarized in Table 2. Further description, additional requirements, clarification and examples of these Protocols follow the table. For an evaluation to be in compliance with the Gross Demand Impact Protocol, the methods used and the way in which data are used and reported must meet all the requirements discussed within this section. The intent is to provide unbiased reliable estimates of program level demand impacts for those programs that are expected to reduce electricity demand. The Protocols sometimes reference the *Evaluation Framework* which provides examples of applicable methods. The requirements,

however, are always those stated in the Protocols, which take precedence over all other protocols and evaluation guidance documents in all circumstances unless otherwise approved or required by the CPUC.

Table 2. Required Protocols for Gross Demand Evaluation

Rigor Level	Minimum Allowable Methods for Gross Demand Evaluation
Basic	Reliance upon secondary data for estimating demand impacts as a function of energy savings. End-use savings load shapes or end-use load shapes from one of the following will be used to estimate demand impacts: 1. End-use savings load shapes, end-use load shapes or allocation factors from simulations conducted for DEER 2. Allocation factors from CEC forecasting models or utility forecasting models with approval through the evaluation plan review process 3. Allocation based on end-use savings load shapes or end-use load shapes from other studies for related programs/similar markets with approval through the evaluation plan review process
Enhanced	Primary demand impact data must be collected during the peak hour during the peak month for each utility system peak. Estimation of demand impact estimates based on these data is required. If the methodology and data used can readily provide 8,760-hour output, these should also be provided.²³ Sampling requirements can be met at the program level but reporting must be by climate zone (according to CEC's climate zone classification). 1. If interval or time-of-use consumption data are available for participants through utility bills, these data can be used for regression analysis, accounting for weather, day type and other pertinent change variables, to determine demand impact estimates. Pre- and post-retrofit billing periods must contain peak periods. Requires using power analysis, evaluations of similar programs, and professional judgment to determine sample size requirements for planning the evaluation. Needs to meet the requirements of the Sampling and Uncertainty Protocol. 2. Spot or continuous metering/measurement of peak pre and post-retrofit during the peak hour of the peak month for the utility system peak to be used with full measurement Option B or calibrated engineering model Option D meeting all requirements as provided in the M&V Protocol. Pre-retrofit data must be adjusted for weather and other pertinent change variables. Must meet the Sampling and Uncertainty Protocol with a program target of 10% precision at a 90% confidence level. 3. Experimental design established within the program implementation process, designed to obtain reliable net demand savings based upon differences between energy consumption during peak demand periods between treatment and non-treatment groups from consumption data or spot or continuous metering.²⁴ Sampling must be according to the Sampling and Uncertainty Protocol.

²³ This includes the use of 15-minute interval data or Building Energy Simulation models whose output is 8,760 hourly data.

²⁴ The overall goal of the Impact Protocols is to obtain reliable net energy and demand savings estimates. If the methodology directly estimates net savings at the same or better rigor than the required level of rigor, then a gross savings and participant net impact analysis is not required to be shown separately.

All M&V referred to in the Impact Evaluation Protocol must be planned, conducted and reported according to the M&V Protocol. M&V may be conducted at a higher level of rigor, with more inputs measured or metered, or with greater precision than the minimum shown within the Impact Evaluation Protocol, but not with a lower level of rigor. The M&V Protocol can also be required by the Joint Staff or used by evaluators to enhance other evaluation efforts.

For the purposes of the Gross Demand Impact Protocol, demand impacts must be reported as energy savings estimates for six time periods for each of four months as follows: noon-1 p.m., 1-2 p.m., 2-3 p.m., 3-4 p.m., 4-5 p.m. and 5-6 p.m. for June, July, August and September for each climate zone in which there are program participants. These demand savings are to be estimated using the Typical Meteorological Year from the National Oceanic & Atmospheric Administration (NOAA), the CEC CTZ long-term average weather data, the Administrator's long-term average weather year or the CEC's rolling average weather year.

The Joint Staff may require that specific studies have additional reporting requirements to include reporting at the 8,760-hour level or specific reporting for targeted transmission or distribution areas. These will be decided on a case-by-case basis as part of the work scoping process or during the evaluation planning process. Identification of these requirements and how they will be met will be incorporated into the evaluation plan and will be conducted and reported as approved within the evaluation planning process.

The Gross Demand Impact Protocol has two rigor levels: Basic and Enhanced. The Basic rigor level uses secondary data to allocate gross energy savings to determine demand savings. The Enhanced level requires primary data collection either through field measurement according to the M&V Protocols or using regression analysis of demand or interval consumption data.

Basic Rigor

The Basic rigor level for the Gross Demand Impact Protocol prescribes that at a minimum, on-peak demand savings are estimated based on allocation of gross energy savings through the use of allocation factors, end-use load shapes or end-use savings load shapes. These secondary data can be from DEER, the CEC forecasting model utility end-use load shape data or other prior studies, with those in the latter two categories needing review and approval through the evaluation planning review process.

Enhanced Rigor

The Enhanced rigor level for the Gross Demand Impact Protocol requires primary data from the program participants. This could be interval-metered data, time-of-use (TOU) consumption billing data, from field measurement or from billing demand data. (This latter is only allowable if the issues of when buildings peak versus demand ratchets and peak periods are addressed in the analysis.) Estimation of peak demand savings estimates is required. If the methodology and data used can readily provide 8,760-hour output, these should be provided. Sampling requirements can be met at the program level but reporting must be by climate zone (according to CEC's climate zone classification). The Joint Staff may require a program evaluation to use the Gross Demand Impact Protocol for transmission and distribution (T&D) demand savings as they deem necessary. Demand evaluation requirements and the methods being employed to meet them need to be clear in the evaluation plans and agreed upon through the evaluation planning review process.

A regression model specified to measure program impacts for peak time periods (via analysis of interval data) or TOU/demand²⁵ consumption metering can be used to estimate program gross demand. This regression analysis must properly account for weather influences that are specific to the demand estimation and other pertinent change variables (e.g., day-type and hours of occupancy). Regression analysis with interval data should focus on obtaining direct demand impacts. If demand consumption data are used, a methodology to estimate demand savings based upon the demand regression analysis must be detailed in the evaluation plan and approved through the evaluation planning review process. Pre- and post-retrofit billing periods must contain peak periods within this analysis. A power analysis in combination with evaluations of similar program and professional judgment must be used to select and justify the proposed sample sizes.²⁶ The evaluation planning, analysis and reporting must meet the requirements of the Sampling and Uncertainty Protocol.

The second class of primary data collection for the Enhanced Gross Demand Impact Protocol is to conduct field measurement of peak impacts within the evaluation effort. Spot or continuous metering/measurement at peak pre- and post-retrofit will be conducted during the peak hour in the peak month for the utility system peak. These data will be used with one of two engineering modeling approaches: (1) full measurement Option B or (2) calibrated engineering model Option D, where the modeling approach must meet all requirements as provided in the M&V Protocol. An overview of the full measurement Option B method is provided in the *Evaluation Framework*.²⁷ The calibrated engineering model Option D method is described and a few references are presented in the *Evaluation Framework*.²⁸ Further information and the specific requirements for the Protocols are provided in the M&V Protocol. Both of these engineering methods need to be designed to a program target of 10 percent precision at a 90 percent confidence level and must meet the requirements of the Sampling and Uncertainty Protocol.

The third class of allowable methods is experimental design with primary data collection. Experimental design with demand measurement comparisons between customers randomly assigned to the treatment and non-treatment groups meets the Enhanced Gross Demand Protocol rigor level. Experimental design will need to measure energy savings during peak periods either through interval data or spot or continuous monitoring of comparison treatment and non-treatment groups to calculate demand savings estimates. Currently, experimental design has not been widely used within efficiency evaluation. The *Evaluation Framework* provides a description and some examples of potential experimental designs within energy efficiency efforts.²⁹ Sampling conducted as part of the experimental design must be conducted according to the Sampling and Uncertainty Protocol.

²⁵ If demand billing is used, the research design must address the issues of building demand versus time period for peak and issues with demand ratchets and how the evaluation can reliably provide demand savings estimates.

²⁶ Power analysis is a statistical technique that can be used (among other things) to determine sample size requirements to ensure statistical significance can be found. There are several software packages and calculation Web sites that conduct the power analysis calculation. Power analysis is only being required in the Protocol for determining required sample sizes. One of many possible references includes: Cohen, Jacob (1989) *Statistical Power Analysis for the Behavioral Sciences*, Lawrence Erlbaum Associates, Inc. Appendix D provides further detail on using power analysis for developing sample size requirements.

²⁷ TecMarket Works, 166-169.

²⁸ Ibid, 129-133 and 176-181.

²⁹ Ibid, 104-105.

Participant Net Impact Protocol

The Participant Net Impact Protocols are summarized in Table 3. Further description, additional requirements, clarification and examples of these Protocols are presented below the table. Being in compliance with the Participant Net Protocol means that the methods used, and the way in which they are used and reported, meet all the requirements discussed within this section. The intent is to provide reliable estimates of program level net energy and demand impacts when combined with the results from work complying with the Gross Energy Impact Protocol and the Gross Demand Impact Protocol. The Protocols sometimes reference the *Evaluation Framework* which provides examples of applicable methods. The requirements, however, are always those stated in the Protocols, which take precedence over all other protocols and evaluation guidance documents in all circumstances, unless otherwise approved or required by the CPUC.

All M&V referred to in the Impact Evaluation Protocol must be planned, conducted and reported according to the M&V Protocol. M&V may be conducted at a higher level of rigor, with more inputs measured or metered, or with greater precision than the minimum shown within the Impact Evaluation Protocol, but not with a lower level of rigor. The M&V Protocol can also be required by the Joint Staff or used by evaluators to enhance other evaluation efforts.

Table 3. Required Protocols for Participant Net Impact Evaluation

Rigor Level	Minimum Allowable Methods for Participant Net Impact Evaluation
Basic	1. Participant self-report.
Standard	- Participant and non-participant analysis of utility consumption data that addresses the issue of self-selection. - Enhanced self-report method using other data sources relevant to the decision to install/adopt. These could include, for example, record/business policy and paper review, examination of other similar decisions, interviews with multiple actors at end-user, interviews with mid-stream and upstream market actors, Title 24 review of typically built buildings by builders and/or stocking practices. - Econometric or discrete choice³⁰ with participant and non-participant comparison addressing the issue of self-selection.
Enhanced	1. "Triangulation" using more than one of the methods in the Standard Rigor Level. This must include analysis and justification for the method for deriving the triangulation estimate from the estimates obtained.

All participant net impact analysis must be designed to estimate the proportion of savings that is program-induced and net of free-ridership estimates (not including spillover savings estimates). This means that it is net of what would have occurred in the absence of the program. The degree to which the research design, selected method, survey instrument design, question wording and model specification can reliably capture this underlying construct is the evaluation's construct

³⁰ The instrumental-decomposition (ID) method described and referenced in the *Evaluation Framework* (page 145) is an allowable method that falls into this category. A propensity score methodology is also an allowable method in this category as described in: Itzhak Yanovitzky, Elaine Zanutto and Robert Hornik,, "Estimating causal effects of public health education campaigns using propensity score methodology." *Evaluation and Program Planning* 28 (2005): 209–220.

validity.³¹ These elements must work together and must be justified based upon how well they address construct validity.

Participant net impact analysis must address the following issues:

Probability that the participant would have adopted the technology or behavior in the absence of the program (participant free-ridership);

- If adopted in the absence of the program, the probability or proportion (partial free-ridership) of expected savings induced by the program given its ability to:
 - Increase the efficiency of what would have been adopted;
 - Make the adoption occur earlier than when it would have occurred; and
 - Increase the quantity of efficient equipment that would have been adopted.
- The estimation of participant net is consistent with decision-making behavior;
- Consistency is assessed to ensure that other forms of bias, such as, centrality bias, are not introduced;
- If survey methods are used, ensuring that survey questions (instrumentation) and techniques are employed to minimize social desirability bias;
- Results that include only free-ridership adjustment are clearly labeled as such;
- Report participant free-ridership and participant spillover separately where the methodologies selected allow this to be done;
- If at least some portion of participant spillover may be embedded within the gross savings estimates cannot be separated out using the estimation method chosen (e.g., a regression approach is used and the spillover behavior is simultaneous with program participation), clearly present why participant spillover may be present within these estimates and a qualitative assessment of whether these might be expected to be significant or not compared to the program savings estimate; and
- If only participant free-ridership is presented in the report without a reporting of participant spillover savings, clearly discuss that this presents a downwardly biased presentation of overall true net savings.

A general discussion of the net-to-gross principals, methods and a few references are presented in the *Evaluation Framework*.³²

The research design, selected method, survey instrument design or modeling specification(s) must also address participant self-selection bias(es). Overall sample size targets can be by program. However, all survey or interview inquiries concerning participant net (free-ridership and spillover, and application to gross impacts to obtain net savings) need to be conducted and measured by measure or end-use. Considerations of uncertainty should guide the sample stratification plan.

³¹ Construct validity refers to the extent to which the operating variable/instrument/survey question accurately taps and properly measures the underlying concept/abstract idea that is designed to be measured.

³² TecMarket Works, 133-146.

Basic Rigor – Self Reports

Participant self-reports is the minimum allowable Basic rigor level method in the Participant Net Impact Protocol. The development of the survey instrument, scoring for responses, and handling of missing data and inconsistent responses needs to address those issues presented above and according to the Sampling and Uncertainty Protocol. A discussion of these issues can be found in the *Evaluation Framework*.³³

Like the other approaches to estimating the net-to-gross ratio (NTGR), there is no precision target when using the self-report method. However, unlike the estimation of the required sample sizes when using the regression and discrete choice approaches to estimating the NTGR, the self-report approach poses a unique set of challenges to estimating required sample sizes. These challenges stem from the fact that the self-report methods for estimating free-ridership involve greater issues with construct validity and often include a variety of layered measurements involving the collection of both qualitative and quantitative data from various actors involved in the decision to install the efficient equipment. Such a situation makes it difficult to arrive at a prior estimate of the expected variance needed to estimate the sample size.

Thus, in order to ensure consistency and comparability, and eliminate potential gaming, this Protocol establishes a minimum sample size for the participant self-report method of 300 participant decision-makers for at least 300 participant sites (where decision-makers may cover more than one site) or a census attempt, whichever is smaller.³⁴ An estimate of the achieved precision for net savings must be reported as well as a detailed description of the method used for its estimation.

Standard Rigor

There are three classes of allowable methods to meet the minimum requirement for Participant Net Impact Protocol Standard rigor level.

Participant / Non-participant Comparison

The first of these is a comparison of participant and non-participant energy consumption that addresses participant self-selection bias. Some of the potential methods to be used are described in the *Evaluation Framework*.³⁵ The evaluation plan and report need to include an analysis and explanation of why the selected research design, methodology and actual model specification were selected. A power analysis in combination with evaluations of similar program and professional judgment must be used to select and justify the proposed sample sizes.³⁶

³³ Ibid, 136-140.

³⁴ This is considered the best feasible approach at the time of the creation of this Protocol. Alternative proposals and the support and justifications that address all of the issues discussed here on the aggregation of variance for the proposed self-report method may be submitted to Joint Staff as an additional option (but not instead of the Protocol requirements) in impact evaluation RFPs and in Evaluation Plans. Joint Staff may elect to approve an Evaluation Plan with a well justified alternative.

³⁵ TecMarket Works, 142-145.

³⁶ Power analysis is a statistical technique to determine sample size requirements to ensure statistical significance can be found. There are several software packages and calculation Web sites that conduct the power analysis calculation. See the Sampling and Uncertainty Protocol for more discussion and reference. Power analysis is only being required in the Protocol for determining required sample sizes. Appendix D provides further detail on using power analysis for developing sample size requirements.

Program-Specific Enhanced Self Reports

The second allowable method is a program-specific enhanced self-report one that draws upon multiple data sources concerning the decision to install/adopt. These could include, for example, record/business policy and paper review, examination of other similar decisions, interviews with multiple actors at the end-user, interviews with mid-stream and upstream market actors, Title 24 review of typical buildings by builders and stocking practices. For commercial/industrial entities multiple decision makers within a firm/corporation could be interviewed, as well as reviews of records and policy documents, and inquiries into decision-making. It also could draw upon either primary data collection or secondary data collection if available on the same California market (from market assessment studies or market effects studies recently completed). The enhanced method could also include engineering components to assist in determining what would have occurred in the absence of the program. Data collected from such multiple sources would be used to triangulate on an estimate of the participant free-ridership and spillover rate for that program. A brief discussion of some of these types of methods and examples is provided in the *Evaluation Framework*.³⁷

Like the other approaches to estimating the NTGR, there is no precision target when using the self-report method. However, unlike the estimation of the required sample sizes when using the regression and discrete choice approaches, the self-report approach poses a unique set of challenges to estimating required sample sizes. These challenges stem from the fact that the self-report methods for estimating free-ridership involve greater issues with construct validity and often include a variety of layered measurements involving the collection of both qualitative and quantitative data from various actors involved in the decision to install the efficient equipment. Such a situation makes it difficult to arrive at a prior estimate of the expected variance needed to estimate the sample size. This Protocol, instead, establishes a minimum sample size for end-use participants: a sample of 300 participant decision-makers for at least 300 participant sites (where decision-makers may cover more than one site) or a census attempt, whichever is smaller. Sample sizes of other actors, engineering work or record review need to be described in the evaluation plan and approved through the evaluation planning review process.

Econometric or Discrete-Choice Analysis

The third allowable method in the Standard rigor level is econometric or discrete-choice analysis of participant and non-participants that addresses participant self-selection bias. An overview of some of these methods and a few references can be found in the *Evaluation Framework*.³⁸ The evaluation plan and report need to include an analysis and explanation of why the selected research design, methodology and actual model specification were selected. A power analysis in combination with evaluations of similar programs and professional judgment must be used to select and justify the proposed sample sizes.

Two of the Standard rigor level methods require comparisons with non-participants. It is important that care be taken for selecting the appropriate comparison group. There is not a single rule about what constitutes an appropriate comparison group, since the selection of the group depends on such factors as type of market transaction, methodology or comparison

³⁷ TecMarket Works, 141-142.

³⁸ Ibid, 142-145.

purpose. Yet, this should be carefully considered and the proposed non-participant comparison group and the criteria used in selecting this group should be discussed in the evaluation plan, and reviewed and approved through the evaluation planning review process.

Enhanced Rigor – Comparison of Multiple Approaches

One of the primary concerns with measurements of participant net is of construct validity. Given this, the Enhanced rigor level requires the use of at least two of the Standard rigor level methods to triangulate³⁹ on an estimate of participant net. This must include analysis and justification for the method for deriving the triangulation estimate, not solely on averages, from the estimates obtained.

Participant net savings evaluation includes the evaluation of free-ridership and participant spillover. Presenting both yields a more accurate picture of what the participant would have done in the absence of the program and the full impacts of the program. The evaluation plan, analysis and report must address how the methods were selected and how the analysis was conducted. Net of free-ridership (Net of FR) estimates must be provided in the evaluation report. Current CPUC policy, as the Protocols are being developed, is that only program savings and not participant spillover will be counted towards program and Administrator goals and performance. These are the Net of FR estimates.

Indirect Impact Evaluation Protocol

The Indirect Impact Evaluation Protocol is the minimum standard Protocol for programs that seek to change the behavior of consumers and for which some level of gross energy and demand savings is expected. These programs are typically information, education, marketing, promotion, outreach or other types that may not have specified energy savings goals, but are still expected to provide energy impacts within their target markets. The Protocol has multiple levels of rigor that can be used to conduct the evaluations. Once a minimum rigor level is assigned for an evaluation, all methods in a higher class of rigor are allowable, as they exceed the minimum criteria. For example, if the program has an assigned the Standard rigor level and the method selected for implementation is an option under the Enhanced rigor level (but is not listed under the Standard rigor level), this method is acceptable for meeting the Protocol.

The Indirect Impact Evaluation Protocol is summarized in Table 4. A discussion of behavioral impact evaluation and selected references are provided in the *Evaluation Framework*.⁴⁰ Further description, additional requirements, clarification and examples of this Protocol follow the table. In order to comply with the Indirect Impact Evaluation Protocol the methods used and the way in which they are used and reported must meet all the requirements discussed within this section. The intent is to provide reliable estimates of program level impacts and, when required, gross energy and demand impacts. The Protocols sometimes reference the *Evaluation Framework* and

³⁹ A strict dictionary definition of triangulation would mean incorporating three measurements. The term is used here to mean a process of analysis that examines at least two measurements and assesses what their differences might mean. Then the best estimate derived from this exam is determined to properly represent the underlying construct to meet construct validity issues to obtain the most reliable estimate from the multiple analyses conducted.

⁴⁰ TecMarket Works, 234-242.

other documents which provide examples of applicable methods. The requirements, however, are always those stated in the Protocols, which take precedence over all other protocols and evaluation guidance documents in all circumstances, unless otherwise approved or required by the CPUC.

All M&V referred to in the Impact Evaluation Protocol must be planned, conducted and reported according to the M&V Protocol. M&V may be conducted at a higher level of rigor, with more inputs measured or metered, or with greater precision than the minimum shown within the Impact Evaluation Protocol, but not with a lower level of rigor. The M&V Protocol can also be required by the Joint Staff or used by evaluators to enhance other evaluation efforts.

Table 4. Required Protocols for Indirect Impact Evaluation

Rigor Level	Minimum Allowable Methods for Indirect Impact Evaluation
Basic	An evaluation to estimate the program's net changes on the behavior of the participants is required; the impact of the program on participant behavior.
Standard	A two-stage analysis is required that will produce energy and demand savings. The first stage is to conduct an evaluation to estimate the program's net changes on the behavior of the participants/targeted-customers. The second is to link the behaviors identified to estimates of energy and demand savings based upon prior studies (as approved through the evaluation planning or evaluation review process).
Enhanced	A three-stage analysis is required that will produce energy and demand savings. The first stage is to conduct an evaluation to estimate the program's net impact on the behavior changes of the participants. The second stage is to link the behavioral changes to estimates of energy and demand savings based upon prior studies (as approved through the evaluation planning or evaluation review process). The third stage is to conduct field observation/testing to <i>verify</i> that the occurrence of the level of net behavioral changes.

Basic Rigor

In this Protocol, programs or program components are assigned by Joint Staff to receive an Indirect Impact Evaluation if the program's primary goal is to produce behavioral changes. The primary uncertainty within the logic chain of obtaining energy and demand savings from these types of programs is the estimation of the program-induced impact on the behavior of participants. Therefore, the primary focus of the Indirect Impact Evaluation is in evaluating and estimating the program's net impact on behavioral change. This is the primary component for the evaluation research design.

There are several types of research design that could be used for conducting an Indirect Impact Evaluation. There are many social science methodologies that could apply depending upon the program goals, logic, program design and market operation. Guidance for these types of evaluations can be found in the *Evaluation Framework*.⁴¹

Indirect impact evaluation design, analysis and reporting must address the following issues:

⁴¹ Ibid, 234-242. Much of the guidance provided from the *Evaluation Framework* chapter on Market Transformation Evaluation (pages 245-268) can also provide useful insights and references.

- Expected impacts and the target audience for these impacts;
- How the expected impacts will be measured;
- Identification and measurement of baseline (and where baseline would have been in the absence of the program, i.e., forecasted, dynamic baseline or estimated counter-factual from research design) or identification and measurement of well-matched non-treatment comparison group over time;
- Extent of exposure/treatment and how this is being measured in the evaluation; and
- Self-selection bias and how this is being controlled for to obtain an unbiased estimate of the program-induced impact.

The assessment or development of a program theory and logic model (PT/LM) is recommended.⁴² The PT/LM could be particularly useful if expanded to include the expected interactions with the market or the use of behavioral change models. These can be valuable as a foundation for the evaluation research design, researchable questions and basis for developing survey/interview questions. Though a PT/LM is not required, it is an important tool to ensure that the evaluation research design can measure the program's behavioral impacts. A detailed evaluation research design for the Indirect Impact Evaluation is required and must be reviewed and approved through the evaluation planning review process.

All sampling must be done in accordance with the Sampling and Uncertainty Protocol. Any sampling for regression analysis must use power analysis in combination with evaluations of similar program and professional judgment to determine required sample sizes.

Methodologies using a treatment/non-treatment group comparison that include controlling for self-selection are encouraged. Methods could include the enhancement of those methods described in the *Evaluation Framework*.⁴³ There are also many other methods used in other evaluation fields that could be found to be equally or more valid. One possible example is the use of the propensity scoring method that has been used to evaluate public health campaigns and control for the selectivity bias in treatment levels.⁴⁴ The evaluation plan and report need to include an analysis and explanation of why the research design, methodology and actual model specification were selected.

Standard and Enhanced Rigor Levels

In the Standard and Enhanced rigor levels, evaluation studies are conducted to link net behavioral impacts to energy and demand saving impacts based upon prior studies. These prior studies do not need to be previously completed evaluations (however this is preferred if they are available). For example, linking net behavior change savings estimates using DEER will meet the Indirect Impact Evaluation Protocol. Linking savings estimates to past evaluations of similar programs, new engineering models for savings estimates or other studies must be approved by the Joint Staff through the evaluation review process.

⁴² Ibid, 30-38 and 45-48.

⁴³ Ibid, 142-145.

⁴⁴ Yanovitzky, Zanutto and Hornik, 209 – 220. Also included (with additional references) in a review of possible net/causality methods for energy efficiency evaluation in: Lisa Skumatz, Dan Violette, and Rose Woods, "Successful Techniques for Identifying, Measuring, and Attributing Causality in Efficiency and Transformation Programs." *Proceedings of the 2004 ACEEE Summer Study in Buildings* (2004): 2.260 – 2.273.

A behavioral impact program (through information, education, training, advertising or other non-monetary incentive efforts) may be part of a portfolio to lead customer/market actors into other programs. This program/program component could be assigned an Indirect Impact Evaluation to determine the impact the program(s) is having on the portfolio and to provide input for the process evaluation of the program. An assignment of the Standard rigor level requires that an impact evaluation be conducted and linked to energy and demand savings estimates. The energy and demand savings, however, would not, in this case, be added to the portfolio level savings unless a method is used and approved by the Joint Staff to ensure that these savings are not double counted with those attributed to other programs.

Four types of impacts from a behavioral change program are shown in Figure 5. Inducing customers into other programs is shown as Path A. Savings from this path are not direct savings due to the information, education, training or advertising program under study. The savings are those obtained through the direct program. However, documenting the impacts of this effort is important to estimate the various components that contribute to generating a portfolio's savings and to aid in making investment decisions. An example might be customers who participate and obtain high-efficiency room air conditioners through a rebate program due to behavioral impacts from the program being evaluated.

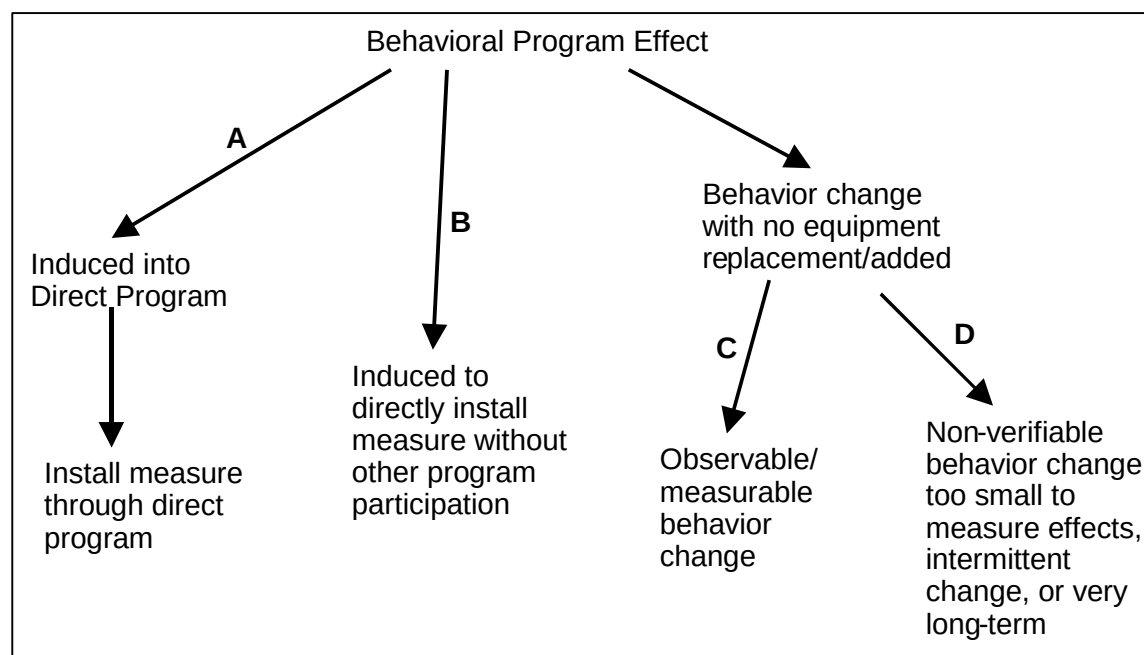


Figure 5. Potential Alternative Behavioral Impact Paths

Programs or program components that directly influence customer behavior to purchase high efficiency replacement equipment or add equipment that can save energy (e.g., timers) are shown as Path B. If assigned an Indirect Impact Evaluation with a Standard or Enhanced rigor level, these programs would be expected to undertake similar evaluation designs to those in Path A. The energy and demand savings for these, however, are *directly* attributable to the program effort being evaluated. The research design may need to estimate and find the proportion of customers

that take these actions outside of other programs. An example might be customers who purchase high efficiency room air-conditioning due only to this program and who did not receive any financial incentives from other portfolio efforts to do so.

Path C refers to those program-induced behavioral changes that can be observed or measured but are not tied to equipment replacement or the addition of equipment. This could include such changes as those to business policies regarding energy efficiency, architects' decisions on when to test daylighting alternatives, and/or plant managers' operating and maintenance schedules.

Path D represents behavioral changes that are too small, long-term or intermittent to be cost-effectively verified through observation, field-testing or surveying with enough reliability to measure any energy and demand impacts. Depending on the level of investment and the advances made in the evaluation of behavioral change, the programs or program components that fall into this category could vary over time. Path D examples include residential behavior of turning off lights, educating children through school programs to changing their energy-use behavior when they are adults, and changes in residential thermostat set points. The Joint Staff will only assign a Basic rigor level for this category if meeting a higher rigor level would not be possible. This could occur because a specific estimate of the degree of the impact cannot be obtained cost-effectively or the link and translation to energy and demand savings is not available or cost-effective to develop.

Every program evaluation is required to demonstrate that the program is accomplishing its primary goals of affecting behavioral change, as stated in its PT/LM.

It is expected that the Indirect Impact Evaluation for paths A, B and C will be assigned either a Standard or Enhanced rigor level depending upon the size of resources being invested and the importance of the anticipated outcomes to the overall success of the portfolio. The indirect impact evaluation for an Enhanced rigor level is distinguished from a Standard rigor level by the requirement to conduct field observation/testing to verify net changes in behavior. For Path D it is expected that only a Basic rigor level will *usually* be assigned. The evaluation design for each path is briefly described below.

Path A: The evaluation design to verify these actions is most straightforward for Path A. Verification through program participation is sufficient given these programs are conducting their own verification and impact evaluation.

Path B: The evaluation design for Path B requires the additional step of finding effected customers. This step would have to be part of the evaluation design when estimating the proportion affected in the impact evaluation. The evaluation plan must propose the research design to accomplish this and be approved within the evaluation planning review process.

Path C: The evaluation research design needed to accomplish an Enhanced rigor indirect impact evaluation following Path C is more challenging. A Path C evaluation plan needs to be presented in enough detail for its logic and potential reliability to be reviewed as part of the evaluation planning review process. Examples of Path C activities include review of pre- and post-program architectural plans, review of government policy, planning and hearing documents

and their dates of adoption along with interview support, examination of business policy manuals, and review of business programs created due to education efforts and testing subsequent employee knowledge and reported actions.

Path D: For path D, the Basic level rigor indirect impact evaluation must be used to demonstrate that the program has carried out specific activities that are designed to produce behavioral change.

Guidance on Skills Required to Conduct Impact Evaluations

The Impact Evaluation Protocol includes gross energy and demand impact Protocols, Protocols for participant net impacts and a Protocol for indirect impact evaluation. There are multiple methods within these various Protocols that create the need for different skills depending upon the method that is being used. The method employed determines the skills and experience requirements for that method. The senior, advisory and leadership personnel for an impact evaluation effort must have the specific skills and experience for the method they are leading and the time budgeted for responsible project task leadership and quality control. The degree of involvement needed from senior skilled staff is dependent upon the skill and experience of the mid-level personnel conducting much of the analysis work.

Several of the energy, demand and participant net methods use statistical/econometric methods. These are used with utility demand metering and consumption data, and with data gathered for decision analysis (in the case of discrete choice). The use of statistical/econometric methods requires personnel trained in these methods and/or with significant experience in using them. This experience and/or training must include testing alternative specifications, testing and correcting for violations of regression assumption violations, and using them within the context of program evaluation.

Another class of methods relies on engineering type methods that draw upon the rules of physics to calculate estimates of energy and demand savings. Simple engineering equations can be understood and used by most people with a general science background. Yet, to ensure reliable use of the principles, impact evaluations using the simple engineering model should still use personnel with experience in this area, Certified Energy Managers⁴⁵ or personnel with training in mechanical or architectural engineering. Building energy simulation models and process engineering models generally require personnel with a college degree in mechanical or architectural engineering or significant related, equivalent experience. Process engineering models may also require specific engineering experience or research regarding the industrial process or facilities being studied.

There are methods within the Gross Energy Impact Protocol (e.g., enhanced gross energy regression-based (enhanced 1.)) that could employ significant primary survey or interview data collection. The participant net impact methods that employ the self-report and enhanced self-report approaches require similar experience and training. The evaluators using these methods should have sufficient experience implementing surveys, interviews, group interviews and other types of primary data collection activities as are being recommended. They need to have

⁴⁵ The Association of Energy Engineers (AEE) offers courses and a certificate for a Certified Energy Manager (CEM).

experience in energy efficiency markets, the social sciences, and interview and survey instrument design, implementation and analysis.

Indirect impact evaluation methods could be based upon survey and interview analysis methods and/or statistical/econometric methods. The evaluators must be trained and experienced in conducting social science research with a strong understanding of assessing and testing causal relationships between exposure to the program and possible outcomes. An important requirement for these evaluators is to have a strong foundation in research design and the ability to create research designs to test for net behavioral impacts of energy efficiency programs.

Summary of Protocol-Driven Impact Evaluation Activities

1	The Joint Staff identifies which programs and program components will receive an impact evaluation and identify the type of impact evaluation(s) to be conducted and at what rigor level.
2	The Joint Staff determines any special needs on a case-by-case basis that will be required from particular program or program component evaluations. CPUC-ED issues request for proposals for impact evaluations, selects evaluation contractors and establishes scope(s) of work.
3	Program theory and logic models (PT/LM), if available, must be reviewed/assessed as needed to properly identify impacts and evaluation elements required to assess net program impacts. Research design and sampling plan developed to meet Protocol requirements at a program or program component basis as designated by the Joint Staff rigor level assignments. This includes meeting requirements from the Sampling and Uncertainty Protocol, M&V Protocol and Reporting Protocol, as are applicable given Impact Evaluation Protocol requirements. Research design and sampling must be designed to meet any of the Joint Staff requirements for additional analyses including, but not limited to, the estimation of net impacts by delivery mechanism, the estimation of transmission and/or distribution benefits, or other areas designated of specific concern by the Joint Staff. Develop Evaluation Plan, submit it to the CPUC-ED and revise as necessary to have an approved Evaluation Plan that meets the Impact Evaluation Protocols.
4	All impact evaluation teams must be staffed so as to meet the skills required for the research design, sampling, appropriate and selected impact evaluation method, uncertainty analysis, and reporting being planned and conducted.
5	Develop precise definitions of participants, non-participants and comparison groups. Obtain concurrence with the CPUC-ED on these definitions which are to be used in developing the research design and sampling plans.
6	All impact evaluations must meet the requirements of the Sampling and Uncertainty Protocol. 6.a There are 2 primary sampling considerations for regression-based consumption analysis. (1) Unless a census is utilized, conduct a power analysis to estimate the required sample size. One may also consider prior evaluations for similar programs and professional judgment (must use all of these for the Enhanced level of rigor); and (2) Must use a minimum of 12 months pre and post-retrofit consumption data, except when program approach does not allow pre-retrofit data (e.g., new construction). 6.b All engineering-based methods must:

	(1) Estimate the uncertainty in all deemed and measured input parameters and consider propagation of error when determining measured quantities and sample sizes to meet the required error tolerance levels; and (2) Use a combination of deemed and measured data sources with sufficient sample sizes designed to meet a 30% error tolerance level in the reported value at a 90% confidence level to meet the Basic rigor level and a 10% error tolerance level at a 90% confidence level for the Enhanced rigor level. 6.c Participant and non-participant comparisons and econometric/discrete-choice methods for Participant Net Impact evaluation will use power analysis combined with examinations of prior evaluation studies for similar programs to derive required sample sizes. 6.d Self-report and Enhanced self-report methods for Participant Net Impact evaluations must at a program level have a minimum sample size of 300 participant decision-makers for at least 300 participant sites (where decision-makers may cover more than one site) or a census attempt, whichever is smaller, (while investigation will be at a measure or end-use level).
7	All impact evaluations must be planned, conducted, analyzed and reported to minimize potential bias in the estimates, justify the methods selected for doing this and report all analysis of potential bias issues as described in the Sampling and Uncertainty Protocol, Impact Evaluation Protocol and M&V Protocol. Primary considerations that must be addressed (based upon method employed) are as follows: 7.a Regression-based consumption analysis must incorporate: (1) Addressing the influence of weather when weather sensitive measures have been included in the program evaluation; (2) Assessing potential bias given inclusion/exclusion issues due to the 12 month pre- and post-retrofit consumption minimum requirement; (3) For the Enhanced rigor level, assess, plan, measure and incorporate background and change variables that might be expected to be correlated with gross and net energy and/or demand savings; (4) Comparison groups must be carefully selected with justification of the criteria for selection of the comparison group and discussion of any potential bias and how the selected comparison group provides the best available minimization of any potential bias; and (5) Interval or TOU consumption data for demand impact analysis must contain the peak period for the utility system peak. If demand billing data is used for demand impact analysis, the research design must address the issues of building demand versus time period for peak and issues with demand ratchets and how the evaluation can reliably provide demand savings estimates. Demand savings must be reported by CTZ. 7.b Engineering-based methods must incorporate: (1) Addressing the influence of weather when weather sensitive measures have been included in the program evaluation; (2) Meeting all the requirements in the M&V Protocol including issues of baseline

	determination; and (3) For the Enhanced rigor level of demand impact analysis using spot or continuous metering/measurement pre- and post-retrofit for the peak hour of the peak month for the utility system peak. Demand savings must be reported by CTZ. 7.c Experimental design must use spot or continuous metering/measurement pre and post-retrofit for the peak hour of the peak month for the utility system peak for determining demand impacts. Demand savings must be reported by CTZ. 7.d Indirect impact analysis must incorporate: (1) Description of expected impacts (direct behavioral and indirect energy and demand impacts) and how they will be measured; (2) Discussion of identification and measurement of baseline; (3) Extent of exposure/treatment and its measurement; (4) Comparison groups must be carefully selected with justification of the criteria for selection of the comparison group and discussion of any potential issues of bias and how the selected comparison group provides the best available minimization of potential bias; and (5) Assessing, planning for and analyzing to control for self-selection bias.
8	Regression analysis of consumption data must address outliers, missing data, weather adjustment, selection bias, background variables, data screens, autocorrelation, truncation, error in measuring variables, model specification and omitted variable error, heteroscedasticity, collinearity and influential data points. These areas must be addressed and reported in accordance with the Sampling and Uncertainty Protocol.
9	Engineering analysis and M&V based methods are required to address sources of uncertainty in parameters, construction of baseline, guarding against measurement error, site selection and non-response bias, engineering model bias, modeler bias, deemed parameter bias, meter bias, sensor placement bias and non-random selection of equipment or circuits to monitor. These areas must be addressed and reported in accordance with the Sampling and Uncertainty Protocol.
10	Develop draft evaluation report to include meeting all requirements in the Reporting Protocol and incorporating the program's performance metrics.
11	Develop final evaluation report in accordance with guidance provided by the Joint Staff. Submit final evaluation report to the CPUC-ED.
12	Once accepted by the CPUC-ED, develop abstracts and post them and report on CALMAC Web site following the CALMAC posting instructions.

Note: The steps included in this evaluation summary table must comply with all the requirements within the Impact Evaluation Protocol.

Measurement and Verification (M&V) Protocol

Introduction

When, in the course of conducting evaluations, it becomes necessary or advisable to collect physical evidence from field installations of energy efficiency technologies, the evaluator must design, document and implement a measurement and verification (M&V) project. M&V will typically be used to support impact studies by providing measured quantitative data from the field. One of the primary uses is to reduce uncertainty in baselines, engineering calculations, equipment performance and operational parameters. However, M&V can be used in process and market effects evaluations as well, when such data are useful for understanding issues such as measure quality and suitability for particular applications, installation practices and quality, baseline equipment efficiency and operation practices, and other issues identified by the process and/or market effects evaluation plan. For the purposes of this Protocol, M&V will cover all field activities dedicated to collecting site engineering information. This includes such activities as measure counts, observations of field conditions, building occupant or operator interviews conducted in-person, measurements of parameters, and metering and monitoring.

How M&V differs from impact evaluation: M&V refers to data collection, monitoring and analysis activities associated with the calculation of gross energy and peak demand savings from individual customer sites or projects. Gross and net impacts at the program level will be guided by the Impact Evaluation Protocol, where results from M&V studies conducted on a sample of sites will be combined with other information to develop an overall estimate of savings by program or program component.⁴⁶

Sources of uncertainty in engineering estimates: Engineering estimates are based on the application of the basic laws of physics to the calculation of energy consumption and energy savings resulting from the implementation of energy-efficient equipment and systems. Engineering models range from simple one-line algorithms to systems of complex engineering equations contained within a building energy simulation program such as DOE-2. Uncertainty in engineering estimates stems from uncertainty in the inputs to an engineering model and the uncertainty in the ability of the algorithms to predict savings.

Uncertainty analysis and M&V planning: Energy efficiency programs utilize a wide range of technical and behavioral tools and concepts as “measures.” The likelihood of success of the measure depends on a large number assumptions, many of which can be verified through measurement. Measured data from field studies are used to quantify and reduce the uncertainty in energy and peak demand impact calculations. While this Protocol is written to support the overall goal of creating more reliable savings estimates and forecasts, we recognize that M&V activities must be planned and resources must be allocated to reduce these uncertainties. Uncertainty analysis conducted during the planning phase shall be used to identify the assumptions that have the greatest contribution to the overall savings uncertainty and allocate resources in an appropriate manner to address these uncertainties.

⁴⁶ It is possible that some impact evaluations will not require M&V. See the Impact Evaluation Protocol herein for more information.

The development of this Protocol is driven by the desire to create and implement a rational framework to identify and conduct a wide range of M&V activities. As the Joint Staff recognizes that precision is a key requirement in forecasting and reporting, it will seek to allocate resources such that the value of the M&V activities is applied to identify, quantify and manage risk associated with the uncertainty in the expected savings from measures and programs. The Protocol supports the overall M&V goals and priorities established by the Joint Staff:

- Improve reliability of savings estimates;
- Determine whether energy and peak demand savings goals have been met;
- Improve DEER estimates of energy and peak demand savings; and
- Inform future program planning and selection processes.

Audience and Responsible Actors

The audience and responsible actors for this Protocol include the following:

- **Joint Staff Evaluation Planners** should understand the uncertainty in the overall energy and peak demand savings calculations and identify the degree to which field measurements can reduce that uncertainty (at appropriate cost);
- **The Evaluation Project Team** will use field measurements to calculate gross savings estimates and answer specific process and market effects evaluation questions;
- **Administrators and IOUs** will use M&V project results to refine unit savings estimates and/or engineering parameters used in future program planning, and utilize early and mid-stream M&V findings to adjust program priorities within the portfolio;
- **Program Implementers** will use early M&V project results to revise program delivery approaches and measures;
- **Site Owners** should allow access to site for field measurements and may have an interest in the energy savings resulting from efficiency upgrades subject to the M&V effort; and
- **DEER Planners** will use field data to develop, calibrate and generally improve DEER energy and demand savings estimates.

Overview of the Protocol

This M&V Protocol is intended to set guidelines for conducting and reporting field data collection activities in support of energy efficiency program evaluations. The M&V Protocol covers the following issues:

- M&V framework;
- Requirements for installation verification;
- M&V requirements;
- M&V approach examples;

- Project reporting and documentation requirements;
- Sampling strategies; and
- Skills required for conducting M&V activities.

For more information on conducting M&V studies, please refer to the *Evaluation Framework*.⁴⁷

M&V Framework & Language

M&V projects conducted under this Protocol shall adhere to the IPMVP,⁴⁸ with additional criteria specified herein. The IPMVP is a flexible framework that allows users to craft M&V plans for specific projects with consideration of:

- The type of contractual arrangement in force;
- The types and quantities of uncertainty in the project savings estimate; and
- The cost to create the M&V plan and conduct all activities in the plan, including:
 - Meter and sensor placement;
 - Data collection; and
 - Data analysis and reporting.

Whereas field measurements are an important component of program impact estimation and the IPMVP is written to allow users flexibility, its application requires a thorough knowledge of measure performance characteristics and data acquisition techniques. Building and energy using facilities in general tend to vary widely in terms of the electrical and mechanical infrastructure that supplies the energy commodity. A measurement strategy that is simple and cheap in one building (such as measuring lighting energy at a main panel) may be much more expensive in a similar building that is wired differently. For this reason, M&V resources, costs and benefits must be called upon and allocated considering site-specific characteristics.

Relationship of the M&V Protocol to Other Protocols

The M&V Protocol is a subset of the Impact Evaluation, Process Evaluation and Market Effects Protocols. M&V activities described within this Protocol are initiated by these three Protocols. Not every evaluation study will require M&V. When M&V is indicated, the M&V Protocol provides the requirements for meeting the various levels of required M&V and points to the applicable pages of the *Evaluation Framework* for more guiding information and references.

Sampling activities conducted within the M&V projects prescribed within this Protocol shall be conducted in accordance with the Sampling Protocol. Impact and process evaluation studies calling for M&V data will include a site selection sampling Protocol.

⁴⁷ TecMarket Works, pages 147-204.

⁴⁸ The IPMVP provides four options for conducting M&V studies. Option C – Whole Facility, is very close in concept to a statistical billing analysis and it is covered under the Impact Evaluation Protocol to avoid confusion.

Key Metrics, Inputs and Outputs

M&V studies, since they are directed by the Impact Evaluation and/or the Process or Market Effects Protocols, will draw upon the same data sources, such as data from program databases, program descriptions, DEER, work papers provided by program implementers, utility demand metering and consumption data for both participants and non-participants, utility weather data, on-site measurement, monitoring and observational data, survey and interview data collection, and other prior study data and reports. These will be used as directed by the M&V Protocol to produce measure-level energy and peak demand savings for sampled sites as directed by the Impact Evaluation Protocol. The overall information inputs and outputs to the M&V process are shown in Figure 6.

Because M&V studies are required to follow the Sampling and Uncertainty Protocol, evaluators must also assess, minimize and mitigate potential bias and present the achieved level of precision including relative precision, error bounds on M&V results in support of the impact evaluation effort.

All M&V reporting must also follow the Reporting Protocol. Verification-only output metrics are defined as the fraction of installed measures that meet the provisions of the M&V Protocol.

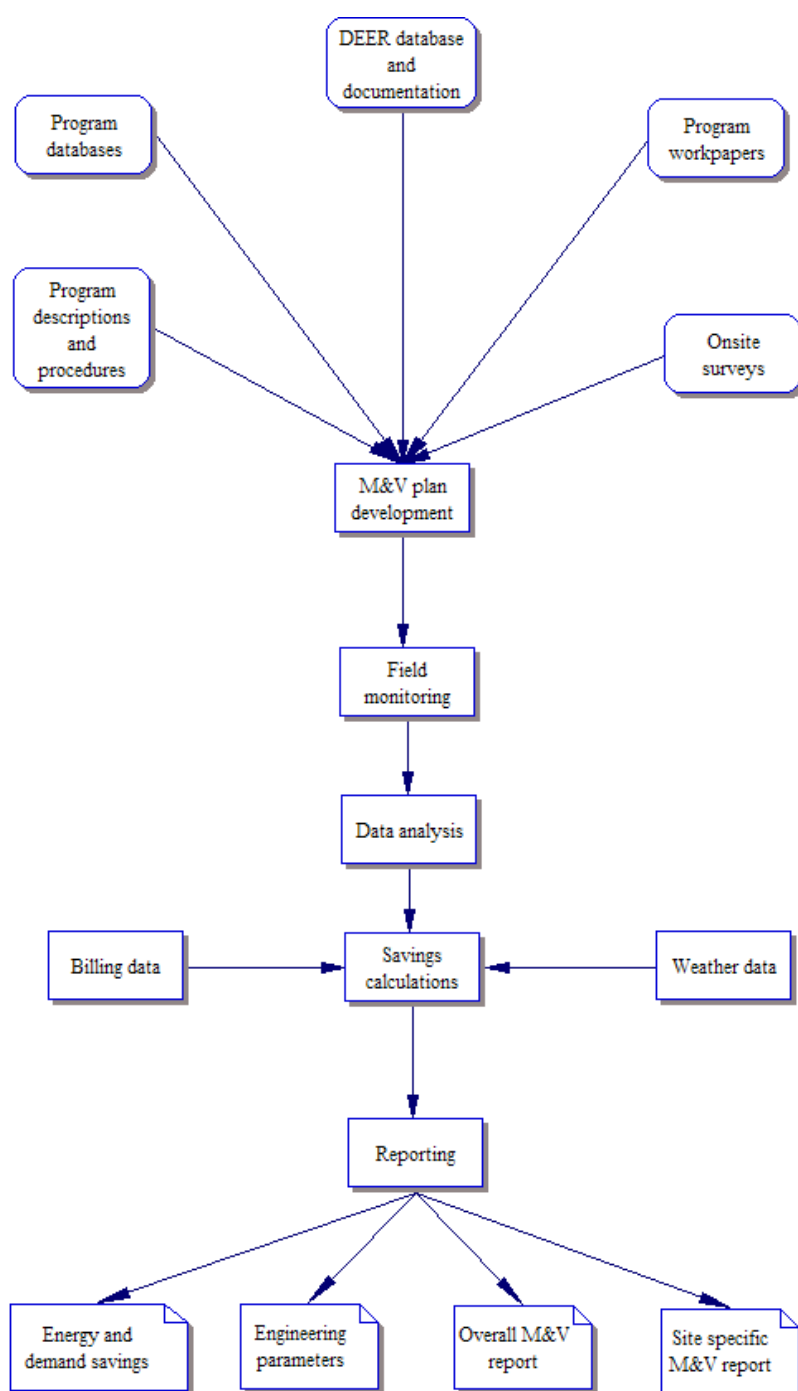


Figure 6. Measurement & Verification Information Flow Diagram

Site-Specific M&V Plan

In requiring the adherence to the IPMVP, this Protocol requires submittal of an M&V plan for each field measurement project undertaken that documents the project procedures and rationale such that the results can be audited for accuracy and repeatability. Within the guidelines established by the IPMVP and the Protocols, there is considerable latitude for the practitioner in developing a site-specific M&V plan and implementing the plan in the field. The M&V contractor shall evaluate the uncertainty in the desired data product and develop a site-specific M&V plan that manages the uncertainty in the most cost-effective manner.

Initial estimates of engineering parameter uncertainties should be used to provide an estimate of the overall uncertainty in the savings calculations. Assumptions used to create initial estimates of parameter uncertainty values should be documented. The contribution of specific engineering parameters to the overall uncertainty in the savings calculations should be identified and used to guide the development of the M&V plan.

The M&V plan must include the following sections:⁴⁹

1. Goals and Objectives;
2. Building Characteristics;
3. Data Products and Project Output;
4. M&V Option;
5. Data Analysis Procedures and Algorithms;
6. Field Monitoring Data Points;
7. Data Product Accuracy (including data acquisition system accuracy and sensor placement issues);
8. Verification and Quality Assurance Procedures (including sensor calibration); and
9. Recording and Data Exchange Format.

The content of each of these sections is described below.

Identify Goals and Objectives: The goals and objectives of the M&V project should be stated explicitly in the M&V plan.

Specify Site Characteristics: Site characteristics should be documented in the plan to help future users of the data understand the context of the monitored data. The site characteristics description should include:

- General building configuration and envelope characteristics, such as building floor area, conditioned floor area, number of building floors, opaque wall area and U-value; window area, U-value and solar heat gain coefficient;
- Building occupant information, such as number of occupants, occupancy schedule, building activities;
- Internal loads, such as lighting power density, appliances, plug and process loads;

⁴⁹ See the *Evaluation Framework*, pages 147-153.

- Type and quantity and nominal efficiency of heating and cooling systems;
- Important HVAC system control set points;
- Changes in building occupancy or operation during the monitoring period that may affect results; and
- Description of the energy conservation measures at the site and their respective projected savings.

Specify Data Products and Project Output: The end products of the M&V activity should be specified. These data products should be referenced to the goals and objectives on the project and include a specification of the data formats and engineering units, with reference to the Reporting Protocol Appendix A.

Specify M&V Option: The M&V option chosen for the project should be specified according to the IPMVP consistent with the M&V Protocol.

Specify Data Analysis Procedures and Algorithms: Engineering equations and stipulated values as applicable shall be identified and referenced within the M&V plan. Documentation supporting baseline assumptions shall be provided.

This is a key component of the M&V plan. Often, data are collected without a clear understanding of the later use for the data. This can result in either extraneous data collection and/or missing data during the data analysis step. Fully specifying the data analysis procedures will help ensure that an efficient and comprehensive M&V plan is presented.

Specify Field Monitoring Data Points: The actual field measurements planned should be specified, including the sensor type, location and engineering units. For example:

- For measuring the run-time of a boiler, the field data point description would be: *“Accumulated run-time of draft fan serving boiler number 1, using an inductive run-time logger mounted on the draft fan motor.”*
- For measuring air conditioner supply air temperature, the field data point description would be: *Duct air temperature (in degrees Fahrenheit) using a sheathed thermistor sensor located in the supply duct three feet downstream from AC-1.*
- For measuring chilled water temperature, the field data point description would be: *“Chilled water supply temperature measured with a probe-type thermistor inserted in a thermowell.”*

Estimate Data Product Accuracy: All measurement systems have error, expressed in terms of the accuracy of the sensor and the recording device. The combined errors should be estimated using a propagation of error analysis and the expected final data product accuracy described.

Specify Verification and Quality Assurance Procedures: Data analysis procedures to identify invalid data and treatment of missing data and/or outliers must be provided.

Specify Recording and Data Exchange Formats: Data formats compliant with the data reporting Protocol should be described.

M&V Rigor Levels

Rigor is defined as the level of expected reliability. The higher the level of rigor, the more confident we are the results of the evaluation are both accurate and precise, i.e., reliable. That is, reliability and rigor are treated as synonymous. Reliability is discussed in the Sampling and Uncertainty Protocol and in the *Evaluation Framework*⁵⁰ where it is noted that sampling precision does not equate to accuracy. Both are important components in reliability, as used by the CPUC.

In accordance with the Impact Evaluation Protocol, M&V requirements are set according to two levels of rigor. The Joint Staff will set rigor levels for each program according to their overall planning priorities as described in the Impact Evaluation Protocol. Each rigor level provides a set of allowable methods that offers flexibility for the M&V contractor to propose the most cost-effective method considering the conditions prevailing at each sampled site. The principle is to establish a minimum level of evaluation rigor. The M&V contractor is free to propose options providing greater rigor than the minimum specified in this Protocol.

Measure Installation Verification

The objectives of measure installation verification are to confirm that the measures were actually installed, the installation meets reasonable quality standards, and the measures are operating correctly and have the potential to generate the predicted savings. Installation verification shall be conducted at all sites claiming energy or peak demand impacts where M&V is conducted. Installation verification activities may also be specified by the Process or Market Effects Protocols.

Measure Existence

Measure existence shall be verified through on-site inspections of facilities. Measure, make and model number data shall be collected and compared to participant program records as applicable. Sampling may be employed at large facilities with numerous measures installed. As-built construction documents may be used to verify measures such as wall insulation where access is difficult or impossible. Spot measurements may be used to supplement visual inspections, such as solar transmission measurements and low-e coating detection instruments to verify the optical properties of windows and glazing systems.

Installation Quality

Measure installation inspections shall note the quality of measure installation, including the level of workmanship employed by installing contractor toward the measure installation and repairs to existing infrastructure affected by measure installation, and physical appearance and attractiveness of the measure in its installed condition. Installation quality guidelines developed by program implementer shall be used to assess installation quality. If such guidelines are not available, they shall be developed by the M&V contractor and approved by the Joint Staff prior

⁵⁰ TecMarket Works, pages 287-314.

to conducting any verification activities. Installation quality shall be determined from the perspective of the customer.

Correct Operation and Potential to Generate Savings

Correct measure application and measure operation shall be observed and compared to project design intent. For example, CFL applications in seldom used areas or occupancy sensors in spaces with frequent occupancy shall be noted during measure verification activities. At enhanced rigor sites, commissioning reports (as applicable) shall be obtained and reviewed to verify proper operation of installed systems. If measures have not been commissioned, measure design intent shall be established from program records and/or construction documents; and functional performance testing shall be conducted to verify operation of systems in accordance with design intent.

M&V Protocol for Basic Level of Rigor

The M&V Protocols for the Basic level of rigor are summarized in Table 5. Further explanations of the provisions of this Protocol follow the table. The M&V contractor is free to propose more rigorous M&V activities during evaluation planning or as directed by the Joint Staff evaluation managers.

Table 5. Summary of M&V Protocol for Basic Level of Rigor

Provision	Requirement
Verification	Physical inspection of installation to verify correct measure installation and installation quality
IPMVP Option	Option A ⁵¹
Source of Stipulated Data	DEER assumptions, program work papers, engineering references, manufacturers catalog data, on-site survey data
Baseline Definition	Consistent with program baseline definition. May include federal or Title 20 appliance standards effective at date of equipment manufacture, Title 24 building standards in effect at time of building permit; existing equipment conditions or common replacement or design practices as defined by the program
Monitoring Strategy and Duration	Spot or short-term measurements depending on measure type
Weather Adjustments	Weather dependent measures: normalize to long-term average weather data as directed by the Impact Evaluation Protocol
Calibration Criteria	Not applicable
Additional Provisions	None

IPMVP Option

The standard M&V Protocol shall conform to IPMVP Option A - Partially Measured Retrofit Isolation.⁵² Savings under Option A are determined by partial field measurement of the energy

⁵¹ Exceptions to this provision are programs offering comprehensive measure packages with significant measure interactions; commissioning, and retrocommissioning programs; and new construction programs. Evaluation of measure savings within these programs conducted using engineering methods must follow the Enhanced rigor M&V Protocol and use building energy simulation modeling under IPMVP Option D.

⁵² See the *Evaluation Framework*, pages 165-166.

use of the system(s) to which an energy conservation measure (ECM) was applied separate from the energy use of the rest of the facility. Measurements may be either short-term or continuous. Partial measurement means that some parameter(s) affecting the building's energy use may be stipulated, if the total impact of possible stipulation error(s) is not significant to the resultant savings. Savings are estimated from engineering calculations based on stipulated values and spot, short-term and/or continuous post-retrofit measurements. Field-verified measure installation counts applied to deemed savings estimates do not meet the requirements of this Protocol.

Sources of Stipulated Data

Stipulated data may be taken from DEER unit energy savings analysis assumptions, efficiency program work-papers, secondary research, engineering references, manufacturers' catalog data, and/or on-site survey data as applicable. Values and sources for stipulated values must be documented in the M&V plan.

Baseline Definition

The baseline used for M&V activities shall be consistent with the baseline definition used by the program. This may include applicable state and/or Federal efficiency standards for appliance or building energy efficiency, existing equipment efficiency or common replacement or design practices as defined by the program evaluated.

Monitoring Strategy and Duration

Spot or short-term measurements may be used, provided the measurement strategy and duration is sufficient to allow calculation of energy and peak demand savings within the uncertainty bounds prescribed by the Impact Evaluation Protocol. Pre-installation monitoring may be required in some cases to meet the applicable uncertainty requirements.⁵³ The *Evaluation Framework* provides more information on monitoring strategy and duration.⁵⁴

Weather Adjustments

Impacts of weather-dependent measures shall be normalized to long-term average weather data as directed by the Impact Evaluation Protocol. Weather conditions prevailing during the monitoring period must be reported. Weather data may be obtained from the nearest representative NOAA or utility weather station or collected on-site. Techniques used to perform the weather adjustments must be documented.

M&V Protocol for Enhanced Level of Rigor

The M&V Protocols for the Enhanced level of rigor are summarized in Table 6. Further explanations of the provisions of this Protocol follow the table. The M&V contractor is free to propose more rigorous M&V activities during evaluation planning or as directed by the Joint Staff evaluation managers.

⁵³ Specific requirements for pre-installation monitoring are not stated in this Protocol, but are a consequence of the uncertainty analysis conducted during M&V planning.

⁵⁴ TecMarket Works, 182-188.

Table 6. Summary of M&V Protocol for Enhanced Level of Rigor

Provision	Requirement
Verification	Physical inspection of installation to verify correct measure installation and installation quality. Review of commissioning reports or functional performance testing to verify correct operation
IPMVP Option	Option B or Option D
Source of Stipulated Data	DEER assumptions, program work papers, engineering references, manufacturers catalog data, on-site survey data
Baseline Definition	Consistent with program baseline definition. May include federal or Title 20 appliance standards effective at date of equipment manufacture, Title 24 building standards in effect at time of building permit; existing equipment conditions or common replacement or design practices as defined by the program
Monitoring Duration	Sufficient to capture all operational modes and seasons
Weather Adjustments	Weather dependent measures: normalize to long-term average weather data as directed by the Impact Evaluation Protocol
Calibration Criteria	Option D building energy simulation models calibrated to monthly billing or interval demand data. Optional calibration to end-use metered data
Additional Provisions	Hourly building energy simulation program compliant with ASHRAE Standard 140-2001

IPMVP Option

The Enhanced rigor M&V Protocol shall conform to IPMVP Option B - Retrofit Isolation⁵⁵ or IPMVP Option D - Calibrated Simulation.⁵⁶ Under Option B, savings are determined by field measurement of the energy use of the systems to which the ECM was applied separate from the energy use of the rest of the facility. Savings are estimated directly from measurements. Stipulated values are not allowed. Under Option D, savings are determined through simulation of the energy use of components or the whole facility. Simulation routines should be demonstrated to adequately model actual energy performance measured in the facility. Savings are estimated from energy use simulation, calibrated with hourly or monthly utility billing data, and/or end-use metering.

Sources of Stipulated Data

Stipulations are not allowed under IPMVP Option B. Under IPMVP Option D, stipulated values used to define the energy simulation model are allowed. Sources of stipulated data may include DEER unit energy savings analysis assumptions, efficiency program work papers, secondary research, engineering references, simulation program default values, manufacturers' catalog data and/or on-site survey data as appropriate. It is impractical to list and reference all data used to define a simulation model. However, model input assumptions that are highly influential in predicting energy and/or peak demand savings shall be identified and documented within the

⁵⁵ See the *Evaluation Framework*, pages 166-168.

⁵⁶ See the *Evaluation Framework*, pages 176-182.

M&V plan. Simulation program name, full version number including applicable release information, and input files shall be provided as documentation.

Baseline Definition

The baseline used for the M&V activities shall be consistent with the baseline definition used by the program. This may include applicable state and/or federal efficiency standards for appliance or building energy efficiency, existing equipment efficiency or common replacement or design practices as defined by the program evaluated.

Monitoring Strategy and Duration

Monitoring shall be sufficient to capture all operational modes and seasons applicable to measure performance. Pre-installation monitoring may be required in some cases to meet the applicable uncertainty requirements.⁵⁷ The *Evaluation Framework* provides more information on monitoring strategy and duration.⁵⁸

Weather Adjustments

Impacts of weather-dependent measures estimated under Option B shall be normalized to long-term average weather data for CEC CTZ in which the site is located. Weather conditions prevailing during the monitoring period must be reported. Weather data may be obtained from the nearest representative NOAA or utility weather station or collected on-site. Techniques used to perform the weather adjustments must be documented. Simulation analysis under Option D shall be conducted using long-term average weather data for CEC CTZ in which the site is located.

Calibration Targets

Building energy simulation models developed under Option D shall be calibrated to monthly energy consumption data. If interval demand data are available, these data shall be used in lieu of monthly energy consumption data. If the modeled floor space area does not match the metered floor space area within ± 20 percent, model calibration is not required. Modelers shall make reasonable attempts to meet the calibration targets listed in Table 7 below. In some cases, forcing a model to meet a particular calibration target may introduce biases in the energy savings estimates. Models not meeting the calibration targets shall be identified and reasons why it is not reasonable to meet these targets must be documented. The Joint Staff may impose additional requirements for short-term end-use monitoring of systems affected by the energy conservation measure during evaluation plan development and review.

Table 7. Model Calibration Targets

Data Interval	Maximum Root Mean Square (RMS) Error	Maximum Mean Bias Error
Monthly	$\pm 15\%$	$\pm 5\%$
Hourly	$\pm 30\%$	$\pm 10\%$

⁵⁷ Specific requirements for pre-installation monitoring are not stated in this Protocol, but are a consequence of the uncertainty analysis conducted during M&V planning.

⁵⁸ TecMarket Works, 182-188.

Additional Provisions

Building energy simulation programs used under Option D shall be compliant with ASHRAE Standard 140-2001.⁵⁹ For example, a partial list of programs compliant with the Standard is shown in Table 8 below:

Table 8. Programs Compliant with ASHRAE Standard 140-2001 (Partial List)

Program	Sector(s)
Micropas	Residential
DOE-2	Residential and Commercial
EnergyPlus	Residential and Commercial

Software using any ASHRAE Standard 140-complaint program as a calculation engine shall be in compliance with this provision of the Protocol.

M&V Approach Examples

This section provides examples of M&V approaches as they apply to specific measure types and rigor levels. The examples are provided for general guidance; M&V contractors are free to proposed M&V plans that are compliant with the Protocols and make sense for the specific site conditions. Example IPMVP options by measure type and rigor level are shown in Table 9 below:

Table 9. Example IPMVP Options by Measure Type and Rigor Level

Measure Type	Basic Rigor Level	Enhanced Rigor Level
Appliances	A	B
Commissioning and O&M programs	D	D
Comprehensive	D	D
Envelope	D	D
Food Service	A	B
HVAC Controls	D	D
HVAC Equipment Efficiency	A	D
Lighting Controls	A	B
Lighting Efficiency	A	B
New Construction	D	D
Non-HVAC Motor Controls	A	B
Non-HVAC Motor Efficiency	A	B
Process	A	B
Refrigeration	A	D
Water Heating	A	B
Water Pumping/Treatment	A	B

⁵⁹ Programs used for non-HVAC simulation, such as industrial processes or refrigeration, do not need to comply with this provision of the Protocol.

Overall Results Reporting

For each M&V project conducted, the M&V contractor must submit a site-specific M&V report. This report is an addendum to the M&V plan submitted prior to conducting field activities and covers the site-specific M&V results and final uncertainty analysis. In addition to the site-specific M&V reports, an overall M&V report shall be filed for each program where M&V activities were conducted within the scope of an individual M&V project contract. The overall M&V report shall include a discussion on the potential sources of bias in the results and steps taken to control and minimize bias, as discussed in the Sampling and Uncertainty Protocol.

Results shall be reported according to the Reporting Protocol and shall conform to the DEER database format as shown in

APPENDIX A. Measure-Level M&V Results Reporting Requirements. Energy and peak demand savings resulting from weather dependent measures shall be reported under weather conditions prevailing during the course of the M&V project. These weather conditions shall be reported along with the energy and peak demand impact information. The impacts shall be normalized to standard weather conditions as directed by the Impact Evaluation Protocol.

Sampling Strategies

M&V projects will be conducted on a sample of program participants and non-participants, as directed by the Impact Evaluation Protocol. Samples drawn for M&V projects shall be congruent with the impact evaluation sample or be nested within the impact evaluation sample where possible. Justification for drawing samples for M&V projects independently from the impact evaluation sample must be provided.

Early scheduling of M&V studies to provide feedback to the program implementer shall be considered in the sample design process. Participant samples for M&V activities may need to be drawn in stages, before the full participant population is established. If problems are identified in early M&V activities and corrected by the implementer, follow-up surveys on a sub-sample of sites may be required to verify that the program delivery modifications are effective.

Samples of measures selected for monitoring at a particular site shall be representative of all measures at the site and shall be selected at random. Measures within a building are often grouped according to similar usage patterns, thus reducing the expected variability in the measured quantity within each usage group. Within each usage group, the sample unit may be the individual measure, a particular circuit or point of control as designated by the M&V plan. Sample units shall be selected at random. Systematic sampling with random start is acceptable. The sampling strategy shall address all measures present at the site that are subject to the M&V study. Target uncertainties for sample designs are specified in the Sampling and Uncertainty Protocol.

Skills Required for M&V

Simple engineering equations and simple instrumentation such as run-time data loggers can be understood and used by people with a general science or engineering background.⁶⁰ Specific training in the use of building energy simulation programs and instrumentation systems is advised but not required.

Summary of Protocol-Driven M&V Activities

1	Receive input from impact evaluation plan. Receive M&V site selection and expected rigor level from the impact evaluation plan.

⁶⁰ The Association of Energy Engineers (AEE) offers a certificate for a Certified Energy Manager (CEM) and a Certified Measurement and Verification professional (CMVP). The material covered in the CEM program is good background for understanding energy engineering concepts addressed by measurement and verification. The CMVP program provides additional training and certification specific to M&V projects.

2	Develop overall M&V plan. The M&V option for each site shall be established according to the rigor assignment and allowable options under the Impact Evaluation Protocol. Project baseline definition with justification shall be reported. Overall M&V planning shall consider the needs of process evaluation studies for measure installation verification and measure performance information. The overall M&V plan shall be submitted for approval to the evaluation project manager as designated by the CPUC-ED.
3	Assess data sources. For each sampled site, the data resources for the engineering analysis must be identified and reviewed. Data sources may include program descriptions, program databases, DEER estimates and underlying documentation, program work papers and on-site surveys. Uncertainties associated with engineering parameters must be estimated. Baseline uncertainties, where not explicitly documented elsewhere, may be informed by professional judgment.
4	Conduct uncertainty analysis. The uncertainty in the estimated savings must be estimated using a propagation of error analysis. The parameters having the greatest influence on the uncertainty must be identified from the propagation of error analysis.
5	Develop site-specific M&V plan according to the outline in the M&V Protocols. The M&V plan must address data collection conducted to reduce uncertainty in the engineering estimates of savings. Sampling of measures within a particular site shall be done in accordance with the Sampling and Uncertainty Protocol. The site-specific M&V plan shall be submitted for review and approval to the evaluation project manager designated by the CPUC-ED prior to commencing field data collection.
6	Conduct pre- and/or post-installation monitoring as indicated by M&V plan. Data collection must be conducted in accordance with the site-specific M&V plan. Changes to the M&V plan resulting from unanticipated field conditions shall be documented and submitted to the evaluation project manager designated by the CPUC-ED.
7	Conduct data analysis and estimate site-specific savings. Conduct analysis of field data and estimate site savings in accordance with site-specific M&V plan. Energy savings estimates for weather-dependent measures shall be normalized to long-term average weather conditions as directed by the Impact Evaluation Protocol.
8	Prepare site-specific M&V report. Prepare a site-specific M&V report for each site used in the analysis that includes the site-specific M&V plan, data collection, data analysis, calculation of measured engineering parameters and overall savings estimates. Calculate the uncertainties associated with energy savings estimates and measurement-derived engineering parameters. The site-specific uncertainty analysis shall include an estimate of the sampling error associated with individual measure sampling within the site, measurement error associated with field data collection and uncertainties associated with any non-measured (deemed) parameters. Potential sources of bias associated with the measurements and engineering analysis shall be identified and steps to minimize the bias shall be reported in accordance with the Sampling and Uncertainty Protocol.
9	Prepare draft overall M&V report. A draft overall M&V project report shall be submitted to the CPUC-ED that meets all the requirements of the Reporting Protocol, demonstrates compliance with the overall M&V plan developed in step 2 and summarizes the results from each site. Site-specific M&V reports shall be included as an Appendix. Raw field data and data analysis results shall be supplied electronically in accordance with the Reporting Protocol.
10	Prepare final overall M&V report. Prepare final overall M&V report in accordance with review comments provided by the Joint Staff.
11	Submit final M&V report. Submit final M&V report and associated datasets to the CPUC-ED.
12	Post final M&V report on the CALMAC Web site. Once accepted by the CPUC-ED, develop abstracts and post them and final M&V report on the CALMAC Web site following the CALMAC posting instructions.

Emerging Technologies Protocol

Introduction

The Statewide Emerging Technologies Program (ETP) is an information-only program that seeks to accelerate the introduction of innovative energy efficient technologies, applications and analytical tools that are not widely adopted in California. The overall objective of the ETP Program is to verify the performance of new energy efficiency innovations which can be transferred directly into the marketplace and/or integrated into utility portfolios in support of resource acquisition goals for energy efficiency. Emerging technologies may include hardware, software, design tools, strategies and services. Finally, it is recognized that such programs are expected to have a number of failures⁶¹ (technologies that do not perform as expected) given the inherent risks⁶² associated with the technologies selected for investigation.

Because of the absence of energy and demand goals and the longer lead time required to introduce new technologies directly into the market and/or into utility energy efficiency programs, a separate Protocol has been prepared to guide the ETP evaluation. The evaluation approach in this Protocol is theory-driven and is based on monitoring the full range of activities, outputs, and immediate, intermediate and long-range outcomes. This approach explicitly recognizes that while many, if not all, of these outputs and outcomes are difficult, if not impossible, to monetize, they can be documented and monitored over time to assess whether the program is on track to achieve the ultimate impacts⁶³.

Because the ETP and other similar programs will evolve over time, the ETP Protocol is designed to be flexible so that the evaluation, measurement, and verification (EM&V) requirements will apply not only to the 2006-2008 ETP but to future ETP designs as well. Of course, the ETP Protocols will also evolve as evaluators gain experience in evaluating such programs.

This Protocol insures a minimum level of evaluation rigor in order to ensure stakeholders that the performance of the emerging technology programs is on-track to achieve their longer-term

⁶¹ There are two types of failure: 1) failure of the technology to perform as expected (note: such failures can provide valuable information to members of the various target audiences), and 2) the failure of the utility to select promising technologies such that a reasonable number of new technologies are not being funneled into utility energy efficiency programs. This Protocol will address both types of failure.

⁶² Risk involves the exposure to a chance of injury or loss (Random House, 1966). Hardware, software, design tools, strategies and services (products) have varying levels of uncertainty as to whether they will perform as expected. Thus, investing in these products assumes varying levels of risk that the return on these investments might not be fully realized (i.e., there will be a loss).

⁶³ Unlike the methods identified in the Impact Protocol, the methods for evaluating the benefits of public investment in RD&D and related emerging technology programs are not nearly as advanced. However, it has been recognized by many that stakeholders should not have to wait three to five to ten years before discovering whether projects with relatively long times are successful (Lee, Russell, Gretchen Jordan, Paul Leiby, Brandon Owens, James Wolf (2003); Link, Albert N. (1996); Ruegg, Rosalie and Irwin Feller (2003); Shipp, Stephanie, Aaron Kirtley, and Shawn McKay (2004); U.S. Department of Commerce, Advanced Technology Program, National Institute of Standards and Technology, Technology Administration (2001); U.S. Department of Commerce, Advanced Technology Program, National Institute of Standards and Technology, Technology Administration. (2001)). There is agreement among many researchers that one should be able to identify immediate and intermediate indicators that can reassure stakeholders that the efforts are on track to achieve such objectives as successful deployment of new technologies into utility energy efficiency programs and the bridging of the "chasm", leading eventually to significant energy and demand impacts.

objectives. This Protocol also provides a wide array of allowable methods in order to offer flexibility for the potential evaluation contractors to propose the most reliable and cost-effective methods that meet the Joint Staff's needs for a given set of evaluation objectives.

Audience and Responsible Actors

The audience and responsible actors for this Protocol include the following:

- Joint Staff evaluation planners – will use the Protocol (1) as input into the ETP evaluation RFPs, and (2) as background and criteria for use in reviewing ETP evaluation plans, managing the ETP evaluations, and reviewing ETP evaluation reports and results.
- Evaluation project team – will use the Protocol to ensure that their detailed ETP evaluation plan(s) meets the requirements in the Protocol. They will also use the Protocol to double-check that the Protocol requirements have been met as they conduct, complete and report the ETP evaluations.
- Portfolio administrators – will use the Protocol to understand how the ETP evaluation will be conducted and to understand the evaluation data needs to support the ETP evaluation. In addition, the Protocol provides background for the administrator's use to determine when to intervene in the program design and implementation efforts to achieve continued and/or greater efficiency gains.
- Program implementers – will use the Protocol to understand the ETP evaluation that will be conducted on their programs and program components. Often, they will be required to provide data to support the evaluation.
- PIER Program administrators – will use the Protocol to understand the ETP evaluation because the activities of the PIER are linked to the activities of the ETP. In some cases, they may be required to provide data to support the evaluation.

Key Metrics, Inputs, and Outputs

ETP evaluations will rely on both secondary and primary data related to various indicators associated with program inputs (e.g., budgets and staff), outputs (e.g., technical reports, articles published, and software) and outcomes (e.g., change in awareness, reduction of performance uncertainty and an increase in adoption rates in the targeted population). *Secondary data* can include, among others, data from program databases, program descriptions, Emerging Technologies Coordination Council (ETCC) databases, work papers developed during program planning, technical reports, white papers, conference papers, on-site measurement and monitoring, and other prior study data and reports. *Primary data* can include, among others, observational data (e.g., on-site visits to demonstration sites), surveys and in-depth interviews with members of the various target populations as well as those who host a demonstration project. Peer reviews can also be conducted using independent experts. *Energy and demand impacts are not performance indicators for the ETP since it is an information-only Program.* These longer-term energy and demand impacts are more appropriately the focus of impact evaluations which will be conducted for utility resource acquisition and market transformation programs after the “new” ETP technologies are deployed in these programs. A more complete listing of possible indicators is provided later in this Protocol. Finally, which data to collect and

what to report are contingent on the size of the evaluation budget, the indicators identified in the program theory and logic model as being the most important, and the chosen methods.

These data will be used within the ETP Protocol's selected methods, a more detailed sample of which is presented later in this Protocol, and conducted through a Joint Staff approved evaluation plan. Unlike resource acquisition programs which are focused on net energy and demand impacts, the performance of the ETP will be based on the preponderance of evidence associated with the analysis of a relatively large number of diverse indicators.

The actual information included in a given report will vary depending on the methods chosen. The specific information to be reported from each study must be determined by the Joint Staff in close collaboration with the independent evaluator.

Evaluation Planning

Once an independent evaluator is hired, the evaluator must prepare a final detailed evaluation work plan that allocates the study's finite resources to maximize the value and use of the information collected while taking into account the requirements of the ETP Protocol. As part of this plan, the evaluator must specifically address the various sources of potential error that are relevant and explain how the resources allocated to each will mitigate the error⁶⁴. The evaluation should also focus on gathering information on specific project and program goals and expectations early in the program cycle from the administrators so that plans can be made to insure that the necessary data are collected.

When samples are used, the ETP evaluation must follow the Sampling and Uncertainty Protocols. Evaluators must assess, minimize, and mitigate potential bias and present, when relevant, the achieved level of precision (including relative precision, error bounds, coefficient of variations, and standard deviations) for interpreting information. It is expected that the aggregate analysis, described later in this Protocol, of *all* ETP projects must first be conducted in order to inform the sampling plan (e.g., the aggregate analysis should shed some light on useful stratification schemes).

The Joint Staff, and other outside stakeholders as deemed appropriate by the CPUC, will review the evaluation plan submitted and discuss with the independent evaluator any tradeoffs they deem necessary to maximize the reliability of the ETP performance assessment. For example, if surveys are conducted of various target audiences, Joint Staff can decide to increase the sample sizes in order to increase precision, recognizing that other sources of error will receive fewer resources or that additional resources may be required to support the change. Or, Joint Staff can decide to reduce the sample sizes and settle for lower precision in exchange for a greater effort to reduce, for example, non-response bias. In the final plan, evaluation resources will be allocated in a way that is consistent with cost-efficient evaluation, i.e., where evaluation resources are set and allocated at levels that maximize the value received from these resources.

⁶⁴ In the pre-1998 Protocols, there was no requirement to address these sources of error in the research plan. Evaluators only had to describe in the final report whether they had to address these various errors and, if so, what they did to mitigate their effects. See Chapter 12 of the *California Evaluation Framework* for further details.

A Sample of Available ETP Evaluation Methods

One of the goals of the ETP Protocol is to combine progress measures for different types of projects in such a way that provides a meaningful assessment of the effectiveness of the ETP program in reaching portfolio level goals like accelerating the introduction of new technologies into utility energy efficiency programs and/or directly into the marketplace. A review of the evaluation literature reveals a number of approaches that could be applied to the ETP evaluation. The following table lists and briefly discusses a number of these methods.

Table 10. Sample of Available ETP Evaluation Methods

Method	Brief Description	Example of Use
Analytical/conceptual modeling of underlying theory	Investigating underlying concepts and developing models to advance understanding of some aspect of a program, project, or phenomenon.	To describe conceptually the paths through which projects evolve or through which spillover effects may occur and validate the underlying theory.
Survey	Asking multiple parties a uniform set of questions about activities, plans, relationships, accomplishments, value, or other topics, which can be statistically analyzed.	To find out how many members of a given target audience have been informed about a given technology through the dissemination efforts of the ETP.
Case study - descriptive	Using single-case or multiple-case designs with single or multiple units of analysis for investigating in-depth a program or project, a technology, or a facility, describing and explaining how and why developments of interest have occurred.	To recount how a particular joint venture (e.g., between the ETP and a customer who hosts a technology demonstration; between the ETP and a manufacturer) was formed, how parties shared research tasks, and why the collaboration was successful or unsuccessful.
Sociometric and social network analysis	Identifying and studying the structure of relationships by direct observation, survey, and statistical analysis of secondary databases to increase understanding of social/organizational behavior and related economic outcomes.	To learn how projects can be structured to increase the diffusion of resulting knowledge.
Bibliometrics - counts	Tracking the quantity of research outputs.	To find how many publications per applied research dollar a technology assessment generated.
Bibliometrics - citations	Assessing the frequency with which others cite publications or patents and noting who is doing the citing.	To learn the extent and pattern of dissemination of a technology assessment's publications and patents.
Bibliometrics - content analysis	Extracting content information from text using techniques such as co-word analysis, database tomography, and textual data mining, supplemented by visualization techniques.	To identify a project's contribution, and the timing of that contribution, to the evolution of a technology.

Historical tracing	Tracing forward from research to a future outcome or backward from an outcome to precursor contributing developments.	To identify apparent linkages between a ratepayer-funded applied research project and something of significance that happens later or has already occurred.
Expert judgment/Peer Review	Using informed judgments to make assessments.	Experts can be called upon to give their opinions about the technical quality and effectiveness of a technology assessment. The experts generally render their verdict after reviewing written or orally presented evidence.

Source: Adapted from Ruegg and Feller (2003)

Protocols Requirements

There is only one level of rigor for the ETP Protocols which has eight required components.

Verification of Basic Achievements

In their 2006-2008 program implementation plans, each utility has established three basic goals that are framed in terms of:

- achieving a certain number of emerging technology application assessments⁶⁵,
- updating the Emerging Technology Database, and
- conducting a certain number of meetings annually of the Emerging Technologies Coordinating Council.

A straightforward verification of whether each utility has met these goals must be conducted. The 2006-2008 ETP verification should include:

- obtaining all relevant documentation of technology assessments launched during the program period⁶⁶,
- comparing the contents of the Emerging Technology Database before the program period and at the conclusion of the program period, and
- documenting the meetings of the Emerging Technologies Coordinating Council.

Beyond 2008, it is assumed that the utilities will continue have a set of basic goals that are amenable to such simple verification. However, independent evaluators must go beyond the simple verification of whether utilities have achieved these basic goals. The remainder of this Protocol describes those activities that must be conducted as a part of a rigorous and comprehensive evaluation of the ETP.

⁶⁵ The technology application assessments may consist of diverse project types including: feasibility studies, simulation analyses, field demonstrations, controlled environment tests, commercial product development, design methodologies and tool development. Some assessments may take up to four years to complete.

⁶⁶ Evaluation consultant contracts will include confidentiality and non-disclosure agreements to cover applicable documents.

Program Theory and Logic Model

Prior to the identification and quantification of performance indicators, the ETP program theory and logic model must be developed. The *California Evaluation Framework* of June 2004 defines program theory and makes an important distinction between a program theory and a logic model:

A program theory is a presentation of the goals of a program, incorporated with a detailed presentation of the activities that the program will use to accomplish those goals and the identification of the causal relationships between the activities and the program's effects. The program theory describes, in detail, the expected causal relationships between program goals and program activities in a way that allows the reader to understand why the proposed program activities are expected to result in the accomplishment of the program goals. A well-developed program theory can (and should) also describe the barriers that will be overcome in order to accomplish the goals and clearly describe how the program activities are expected to overcome those barriers. A program theory may also indicate (from the developers perspective) what program progress and goal attainment metrics should be tracked in order to assess program effects.

Program theories (PT) are sometimes called the program logic model (LM). A stricter definition would be to differentiate the program theory as the textual description while the logic model is the graphical representation of the program theory showing the flow between activities, their outputs, and subsequent short-term, intermediate, and long-term outcomes. Often the logic model is displayed with these elements in boxes and the causal flow being shown by arrows from one to the others in the program logic. It can also be displayed as a table with the linear relationship presented by the rows in the table. The interactions between activities, outputs, and outcomes are critical to understanding the program logic and argue for the need to have, or construct, both a program theory and a program logic model. (p. 31)

A more thorough discussion of program theory and logic models can be found in Chapter 4 of the *California Evaluation Framework*.

Describing the various ETP activities and how these activities interrelate to produce immediate, intermediate, and long-term outputs and outcomes is a necessary first step. These outputs and outcomes can be considered additional objectives beyond the three basic objectives describe above. Once described, the underlying theory must be explicated, i.e., why are these activities expected to achieve these outputs and outcomes. As part of this process, immediate, intermediate, and long-term indicators of progress toward the ultimate goals will be identified. Some of these indicators are easily quantifiable (number of papers and patents, amount of additional investment) and others are somewhat more difficult to quantify (changes in behavior, changes in procedures). *While the indicators pursued by the independent evaluator should be guided by the logic model, there might be other indicators that the CPUC wishes to pursue that are related to objectives other than those explicitly noted in the logic model.*

As a part of the development of the program logic model, the various target audiences for the ETP activities must be identified. Once the program theory and logic model have been developed, future evaluation efforts must review the logic model and theory to determine if changes are needed. Finally, it is recognized that, while there will be a statewide ETP theory and logic model, it is possible that utility-specific program theories and logic models will be required if each utility's ETP deviates in important ways from the statewide theory and logic model.

Aggregate Level of Analysis

The aggregate analysis is designed to achieve two objectives:

- To describe, for each utility, the basic components or elements that make up the ETP and provide the necessary broader context for assessing the performance of the ETP (e.g., budgets, FTEs, types of technology assessments, average duration of projects, collaboration with other institutions/agencies, etc), and
- To determine, for each utility, the extent to which the overarching program and policy objectives have been met (e.g., addressing the needs of all customer sectors, assuming acceptable levels of risk, etc.).

The aggregate analysis involves the analysis of a variety of data collected for *all* of the projects in each utility's ETP portfolio. Such a level of analysis provides a statistical overview of the ETP portfolio (e.g., frequencies, cross tabulations, means etc.) across multiple projects and participants in order to achieve the two objectives listed above. The analysis of these aggregate data will allow one to address a number of contextual, program and policy questions, such as:

1. What are the various sources of funding, (PGC, academic institutions, manufacturers, government agencies, etc.), by type of technology assessment?
2. How many full-time equivalent ETP employees are involved by type of technology assessment?
3. How does PGC funding and co-funding vary by type of technology assessment by sector over time?
4. How does PGC funding and co-funding vary by end use and/or by sector over time?
5. What is the frequency of the various types of technology assessments, by end use, over time?
6. How is risk being balanced (e.g., measures that do not perform as expected versus those that do)?
7. What is the average duration of a technology assessment?
8. Are the technology assessments proportionately focused on sectors and end uses in which there are the greatest expected potential energy and demand benefits?
9. How many technology assessments are launched annually?
10. How many technology assessments are currently active?
11. What percent of the technologies sponsored by the ETP have been deployed into utility energy efficiency program and/or directly into the marketplace?
12. Are there imbalances in the types of projects funded?
13. Are the needs of all the sectors being adequately addressed?

Data for ETP assessment can be collected using a survey of key ETP staff along with extracts from the program database or ETCC database. *Examples* of data that could be collected for the aggregate analysis include:

1. Funding by the PGC and by other entities (authorized budget, invoiced and committed)
2. Stage of development for each technology
3. Specific technologies and end uses
4. Expected long-term energy and demand benefits from each project (provided by ETP program staff and/or the ETCC database) and the possible timeline of those forecast.
5. Project initiation and completion (date on which all work has ceased) dates
6. Failures (technologies that do not perform as expected based on ETP analysis) as a percent of all projects
7. Subjective assessment of risk
8. Targeted sectors and population(s) within that sector,
9. Whether the technology has been deployed into a utility energy efficiency program and/or directly into the marketplace.

The eventual list of key variables will be determined in close collaboration with the CPUC-ED, the independent evaluator and ETP staff.

Implementation Analysis

The final task is to conduct a program- and utility-specific analysis to determine whether there have been any deviations from the program implementation plan, as described in the program theory and logic model. Any deviations from the plan and implementation problems must be explained. This analysis must focus on such issues as the selection process used by ETP managers to select “promising” projects, collaboration between PIER, the ETP, and utility program staff, and unanticipated problems and their resolution. This analysis must be initiated early in the program period so that any necessary corrective guidance can be provided to program administrators on an on-going basis. *Independent evaluators should look for opportunities to collaborate with utilities, which are responsible for conducting process evaluations of the ETP.*

Measure Tracking

Those technologies that have been deployed to utility energy efficiency programs must be tracked over time to determine their adoption rates⁶⁷ and resulting energy and demand impacts. Adoption rates and energy and demand impacts are useful indicators of how well the ETP screened promising technologies and developed strategies, in close collaboration with the utility-sponsored energy efficiency programs, to cross the “chasm”. The goal of this component of the Protocol is *not* to attribute these savings directly to ETP as a resource, but to show a clear trail of which ETP technologies are being accelerated into utility energy efficiency programs. Only by planning for this type of tracking can an evaluation adequately answer the future questions posed by key stakeholders regarding the ultimate impacts of ETP activities.

⁶⁷ Adoption rates (e.g. the number of measures adopted on an annual basis) for various measures installed through utility resource acquisition programs and associated energy and demand impacts will be obtained from utility program tracking databases. This is generally considered as distinct from a market penetration rate or a saturation rate.

While the previous five components are focused on the entire ETP, including all of the technology assessments, the next three components focus on samples of projects.

Detailed Analysis of Key Performance Indicators

This component involves the collection of additional data that address a number of areas, such as: 1) knowledge creation, 2) knowledge dissemination 3) technical progress, 4) progress towards commercialization, and 5) the deployment of new measures to utility-sponsored energy efficiency programs. Specifying the indicator variables for the ETP should be guided by the ETP logic model, which identifies short, intermediate, and long-term outcomes associated with diverse projects. Some *examples* of project-level indicators for which data could be collected are:

- knowledge created
 - technical papers
 - articles published
 - technical reports
 - conference presentations
 - fact sheets
 - brochures
- knowledge disseminated
 - technical reports distributed and to whom
 - number and content of workshops and professional forums
 - conference presentations, topics and dates and estimated size of audience
 - number of fact sheets distributed and to whom
 - brochures distributed and to whom
 - websites created (includes hits on the websites and downloads)
 - bibliometric counts
- number of demonstration projects
- performance data collected at demonstration sites
- technical and market barriers overcome, technical milestones met, and significant knowledge gained
- remaining technical and market barriers
- prototypes developed and prototypes passing performance tests
- patents (both filed and granted)
- licenses
- awards for excellence
- interviews with those hosting the demonstration projects
- collaboration with manufacturers
- the number and description of new measures being deployed directly into the marketplace and/or into utility programs.

Depending on the nature of the project, one could also examine the extent to which the project has attracted capital for advancing commercialization objectives, including resources provided by any funding partners.

Finally, for each selected project, the reasons why it was selected must be discussed in terms of the selection criteria. Such topics as the technology's technical and economic energy and demand potential, description of the targeted populations, the identified risk factors, market barriers, the existence of known delivery channels, and the evidence that there was a need to need for a bridging function could be discussed.

All data must be systematically analyzed so that an overall assessment of each utility's ETP with respect to its specific objectives can be conducted by the independent evaluator. These objectives must be determined early in the program cycle, as part of the development of the logic model, so that a plan to gather the necessary data can be designed.

If there are fewer than 30 projects⁶⁸ within a given utility during the program period, a census of all projects must be conducted. If there are more than 30 projects, then a random sample of projects must be evaluated. The size of the sample must be determined by the independent evaluator in close collaboration with the Joint Staff. The sample design must be informed by the aggregate analysis. In addition, the sample of projects for each utility should be stratified by size of budget, the level of uncertainty regarding success, or the magnitude of expected benefits. The stratification variable will be selected after the aggregate analysis.

This next two components have two objectives: 1) to conduct a more rigorous assessment of the technical achievements of selected ETP projects through the use of a peer review panel⁶⁹, and 2) to provide a more definitive assessment of the extent to which the "chasm", defined as a discontinuity in the product life cycle that occurs from early adopter to the mass market (Moore, 2002)⁷⁰, has been bridged. Projects selected for these next two components should be nested within the sample of those selected for Detailed Analysis of Key Performance Indicators.

Peer Review

A random sample of the ETP projects for each utility must be subject to a technical review using the peer review process. For example, such projects as the laboratory testing of refrigeration measures could be subjected to a technical review in order to evaluate the quality of the research process and output (e.g., whether the design of the study was sound, whether the project provided any new insights on the assessed technology). The focus should be on those projects in the highest strata (i.e., those with the largest budgets, the greatest uncertainty regarding success, or the greatest expected benefits identified in the previous component, *Detailed Analysis of Key Performance Indicators*). The number of projects that are peer reviewed for each utility and the extent of each review must be determined based on the size and complexity of projects and the size of the evaluation budget.

⁶⁸ A project can cover a variety of activities associated with a technology application assessment including feasibility studies, simulation analyses, field demonstrations, controlled environment tests, commercial product development, design methodologies and tool development. Some assessments may take up to four years to complete.

⁶⁹ See the *Peer Review Guide* prepared by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) Peer Review Task Force, 2004.

⁷⁰ The chasm separates the early adopters from the early majority. Crossing the chasm requires that those in the early majority receive something that the early adopters do not need, the needed assurances from trusted sources regarding new technologies. Many new products fail because they are not able to cross the chasm in terms of new product design and marketing strategy, from the early market (early adopter) to the mass market (early majority).

Peer reviewers will be selected by the evaluation contractor in close collaboration with the CPUC-ED with input from the utilities. Each potential reviewer will be asked to identify any areas related to this project where a conflict or appearance of conflict could exist and explain the nature of that conflict. A key resource regarding the use of peer reviewers is the “PEER REVIEW GUIDE: Based on a Survey of Best Practices for In-Progress Peer Review.” This document was prepared in 2004 by the Office of Energy Efficiency and Renewable Energy (EERE) Peer Review Task Force for U.S. Department of Energy, Energy Efficiency and Renewable Energy.

Target Audience Surveys

To assess the extent to which the chasm is being crossed, surveys⁷¹ of members of the various target audiences (end users and those upstream from the end users including those who request materials, download materials, are directly sent materials, visited demonstration sites, and attended conferences and workshops) must be conducted in order to determine the impact of knowledge dissemination on the targeted populations with respect to any reductions in key market barriers and any subsequent increases in the adoption of ETP technologies. Of course, this requires that in the development of the program logic model the various target audiences for the ETP activities must be identified and that baselines are established so that progress can be measured.

Integration of Results

The results for each utility must be aggregated across the projects examined so that, based on the preponderance of the evidence, conclusions regarding a utility's performance with respect to its entire ETP portfolio can be assessed. These results must then be aggregated across utilities so that the performance of the statewide ETP, based on the preponderance of the evidence, can also be assessed. Various approaches to aggregating performance indicators are available including Keeney and Raiffa (1993), Reugg and Feller (2003), and Shipp et al. (2004).

Reporting of Results

The Emerging Technology Program Evaluation will be reported consistent with the requirements for all evaluation reports described in the Reporting Protocol in the section entitled “Common Evaluation Reporting Requirements.” In addition, the following elements should be included in the evaluation reports under the Methods heading.

- Program Theory and Logic Model
- Goal Verification
- Aggregate-Level Analysis
- Implementation Analysis
- Measure Tracking
- Detailed Analysis of Key Performance Indicators
- Peer Review
- Target Audience Surveys

⁷¹ Whenever surveys are based on samples, the *Sampling and Uncertainty Protocols* apply.

These presentations must be provided in enough detail that the differences (if any) in the methodological approach across different technologies and utilities can be understood by the reader. Finally, one must describe the approach for integrating the study results so that the overall performance of the ETP can be assessed.

The Reporting Protocols includes a requirement that all evaluation reports include a presentation of the detailed study findings. This presentation must be provided in enough detail that the different results or findings (if any) can be understood for each technology assessment covered in the study. The report should present the results of each of the required eight components contained in the ETP Protocol. Reports will be provided consistent with the Reporting Protocol.

Summary

The following table provides a summary of the Protocol that can be used to guide the evaluation efforts once the detailed contents of the Protocol are well understood.

Summary of Protocol-Driven Emerging Technology Evaluation Activities

1	Joint staff selects an evaluation contractor to implement the Emerging Technology Program evaluation.
2	The ETP managers, in collaboration with the evaluation contractor and the CPUC-ED, develop logic models and program theories to inform the evaluation plan.
3	The contractor works with the CPUC-ED on the development of the draft evaluation plan (with possible input from the program implementer) consistent with the ETP Protocol. As necessary, the plan must comply with the other Protocols (Impact Evaluation Protocol, Process Evaluation Protocol, Market Effects Protocols, the Sampling and Uncertainty Protocol and the Reporting Protocol) in the development of the evaluation plan and in the implementation and reporting efforts.
4	The CPUC-ED works with the evaluation contractor to finalize and approve an evaluation plan from which the contractor can begin the evaluation effort.
5	The contractor carries out all eight of the required Protocol requirements in order to measures key short, intermediate, and long-range performance indicators identified in the logic model.
6	The contractor reports the results of the final evaluation to the CPUC-ED and Joint Staff consistent with the provisions in the Reporting Protocol.
7	Once the report is accepted by the CPUC-ED, the contractor develops abstracts and posts the report on CALMAC web site following the CALMAC posting instructions.

Note: the steps included in this evaluation summary table must comply with all the requirements within the Emerging Technology Protocol in order to be in compliance. Any deviations from the Protocol must be agreed to by Joint Staff and fully documented within the evaluation plan and in the evaluation report.

References

- Ausubel, Jesse H. and H. Dale Langford, 1997, *Technological Trajectories and the Human Environment*, New York: National Academy Press.
- Boardman, Anthony E., David H. Greenberg, Aidan R. Vining, and David L. Weimer, 1996, *Cost-Benefit Analysis: Concepts and Practice*, Upper Saddle River, N.J.: Prentice Hall.
- Breland, H. (1983). *The direct assessment of writing skill: A measurement review* (Tech. Rep. No. 83-6). Princeton, N.J.: College Entrance Examination Board.
- California Public Utilities Commission, 2001, *California Standard Practice Manual: Economic Analysis of Demand-Side Programs and Projects*.
- Cohen, J.: "A Coefficient of Agreement for Nominal Scales." *Educational Psychological Measurement*, 20, 37-46, 1960.
- Edwards, Ward and J. Robert Newman, 1982, *Multiattribute Evaluation*, Thousand Oaks, CA: Sage Publications.
- Edwards, Ward, Marcia Guttentag and Kurt Snapper, 1975, "A Decision Theoretic Approach to Evaluation Research," In Elmer L. Struening and Marcia Guttentag (Eds.), *Handbook of Evaluation Research*, Beverly Hills, CA: Sage Publications, Inc.
- Fleiss, Joseph L. *Statistical Methods for Rates and Proportions* 2nd Edition. New York: John Wiley & Sons, 1981.
- Law, J. Randolph and Paul J. Sherman. "Do Raters Agree? Assessing Inter-Rater Agreement in the Evaluation of Air Crew Resource Management Skills." In *Proceedings of the Eighth International Symposium on Aviation Psychology*, Columbus, Ohio: Ohio State University, 2001.
- Lee, Russell, Gretchen Jordan, Paul Leiby, Brandon Owens, James Wolf, 2003, "Estimating the Benefits of Government-Sponsored Energy R&D," *Research Evaluation*, Vol. 12 No. 3, pp. 189-195.
- Mahajan, Vijay, and Robert A. Peterson, 1985, *Models for Innovation Diffusion*, Newbury Park, CA: SAGE Publications.
- Mahajan, Vijay, Eitan Muller, and Rajendra K. Srivastava, 1990, "Determination of Adopter Categories by Using Innovation Diffusion Models," *Journal of Marketing Research*, Vol. XXVII, pp. 37-50.
- Mahajan, Vijay, Eitan Muller, and Yoram Wind, 2000, *New-Product Diffusion Models*, Boston: Kluwer Academic Press.

McGraw, K. O. and S. P. Wong. (1966). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, Vol. 1, No. 1, 30-46.

Random House, 1966. The Random House Dictionary of the English Language, ed. Jess Stein. New York: Random House.

RER, 1999, "Suggestions for Incorporating Multiperiod Market Transformation Effects and Energy Center Informational Impacts in the Public Purpose Tests for Some 2000 Programs," Prepared by Regional Economic Research for the Southern California Edison Company.

Ridge, Richard, 2000, "Evaluation of Southern California Edison's Emerging Technology Program," Submitted to the Southern California Edison.

Ridge, Richard, 2003, "Evaluation of the 2002 California Emerging Technologies Program," Submitted to the Southern California Edison.

Rogers, Everett M., 1995, *Diffusion of innovations*, New York: The Free Press.

Rogers, Everett M. and F. Floyd Shoemaker, 1971, *Communication of innovations: A cross-cultural approach*, New York: Free Press.

Ruegg, Rosalie and Irwin Feller, 2003, *A Toolkit for Evaluating Public R&D Investment: Models, Methods, and Findings from ATP's First Decade*. Prepared for the Economic Assessment Office, Advanced Technology Program, National Institute of Standards and Technology, Gaithersburg, MD.

Saaty, T.L., 1980, *The Analytic Hierarchy Process*, New York: McGraw-Hill.

Sebold, Frederick D., Alan Fields, Lisa Skumatz, Shel Feldman, Miriam Goldberg, Ken Keating, and Jane Peters, 2001, *A Framework for Planning and Assessing Publicly Funded Energy Efficiency (Study ID PG&E-SW040)*, Prepared for Pacific Gas & Electric.

Shipp, Stephanie, Aaron Kirtley, and Shawn McKay, 2004, *Evaluating Individual Projects and the Portfolio of Completed Projects in the Advanced Technology Program*. Unpublished manuscript.

Soft, Steven, 2002, *Power System Economics: Designing Market for Electricity*, Piscataway, N.J.: IEEE Press/Wiley-Interscience, John-Wiley & Sons, Inc.

Sultan, Fareena, John U. Farley, and Donald R. Lehmann, 1990, "A meta-analysis of applications of diffusion models," *Journal of Marketing Research*, Vol. XXVII, pp. 70-

"The California Evaluation Framework." (June 2004). Prepared for the California Public Utilities Commission and the Project Advisory Group by TecMarket Works, Megdal & Associates, Architectural Energy Corporation, RLW Analytics, Resource Insight, B & B Resources, Ken Keating and Associates, Ed Vine and Associates, American Council for an Energy Efficient Economy, Ralph Prah and Associates, and Innovologie.

U.S. Department of Commerce, Advanced Technology Program, National Institute of Standards and Technology, Technology Administration. "Performance of 50 Completed ATP Projects: Status Report - Number 2 (NIST Special Publication 950-2)." U.S. Government Printing Office, Washington, D.C., December 2001.

U.S. Department of Commerce, Advanced Technology Program, National Institute of Standards and Technology, Technology Administration. (2001). "Measuring ATP Impact: 2004 Report of Economic Progress." U.S. Government Printing Office, Washington, D.C.

U.S. Department of Energy, Energy Efficiency and Renewable Energy. (2004). "PEER REVIEW GUIDE: Based on a Survey of Best Practices for In-Progress Peer Review." Prepared by the Office of Energy Efficiency and Renewable Energy (EERE) Peer Review Task Force.

Codes and Standards and Compliance Enhancement Evaluation Protocol

Introduction

This Protocol covers approaches for evaluating codes and standards programs, and for evaluating code compliance enhancement programs. The primary focus of this Protocol is to present the approach for documenting savings from the California Codes and Standards Program and the evaluation of Code Compliance Programs yet to be developed and implemented. The Code Compliance Enhancement Protocol is being added at this time because the IOUs are considering the addition of compliance enhancement programs into their energy efficiency program portfolio. The Compliance Enhancement Program Evaluation Protocol is new and has never before been applied within the evaluation community. As a result it is designed to be flexible, allowing a wide range of approaches to be conducted once they are approved by the Joint Staff.

This Protocol describes how gross and net energy savings will be estimated for programs that change or contribute to a change in building codes or appliance standards that are expected to result in energy savings and programs that are implemented to increase the level of compliance with code requirements. It does not cover process evaluations or other types of evaluations that may address additional research goals. Other sections of the Protocols provide instructions on these studies. This Protocol identifies a series of evaluation-related activities that produce estimates of gross and net energy saving from Codes and Standards Programs and net energy savings from Code Compliance Programs. In addition, this Protocol identifies the audience and responsible actors associated with these evaluation efforts, the key metrics to be produced from the evaluations, the change theories and the logic models that need to detail the assumed causal relationships for achieving the savings, and the evaluation approach that is to be used to estimate gross and net program impacts. These issues are discussed below.

We note early in the Protocols that codes and standards evaluations that follow this Protocol are best contracted prior to and launched at the same time that the CEC is assessing which technologies should be considered for the next round of codes or standards changes. This effort is launched approximately three years before a change begins producing energy savings. The evaluations of the Code Compliance Enhancement Programs should be launched at the same time the programs are first launched so that baseline compliance assessment can be compared to post-implementation changes in compliance.

The evaluation contractor selected to conduct the evaluation of the Codes and Standards Programs will need to realize that the change theories and logic models developed by the program will be adjusted and expanded or contracted from time to time as new change-related causal relationships are identified and as program activities are modified to meet the program's objectives. These conditions will require a multi-year evaluation effort that is timed to the code program's change process rather than the program implementation cycles, so that the evaluation contractor can be charged with the responsibility to evaluate a specific set of assigned code or standard changes. As additional code changes are developed over time, additional evaluation contracts will be awarded to cover the code changes not included in the previous group of

evaluated changes. This means that there will be periods of time in which multiple evaluation contracts may be in force to evaluate the program, but these studies will focus on a different set of changes.

The evaluation activities conducted under this Codes and Standards Protocol are established to be both prospective and retrospective. They are designed to assess events and conditions that occur in the future, such as the projected energy savings to be achieved. However, they are also designed to be retrospective, with true-up efforts that look back over time and adjust evaluation findings to reflect actual market conditions. As such the evaluations may be contracted in two phases, with the first phase being the assessment and projection of current and future savings, followed by true-up studies that look back and adjust the projected findings and energy savings to reflect actual construction, retrofit, and purchase patterns.

The evaluations conducted under the Codes and Standards Protocol will need to be staffed and managed to be adaptive to the different stages associated with the different activities of the change process that will occur at different times. The evaluation contractor must be aware that they will need to coordinate with the program administrators to be able to respond to the different efforts and activities with the right evaluation activities at the right time.

Finally, Both the Codes and Standards Protocol and the Compliance Enhancement Protocol included at the end of the Protocol is new to the evaluation industry. As they are used and tested over the next few program cycles it will need to be updated to reflect the experiences of the first sets of evaluations conducted under these Protocols. Likewise, all Protocols need to be updated periodically as new methods and approaches are developed and as the evaluation reporting needs change.

Audience and Responsible Actors

The audience and responsible actors for this Protocol include the following:

- Joint Staff Evaluation Planners – will use the Protocol to develop evaluation RFPs for the impact evaluation contracts to review and supervise the evaluation contractors to assure adherence to the Protocol, to describe the evaluation's focus and approach to the evaluation stakeholders and information consumers, and to meet other needs identified by the Joint Staff.
- Evaluation Contractors – will use the Protocol to develop their detailed evaluation plan in accordance with Protocol requirements and provide unbiased, objective, and independent evaluation results. They will use the Protocol to guide the evaluation effort and to ensure that the Protocol requirements have been met and the evaluation report provides the required information.
- Portfolio Administrators – will use the Protocol to understand how the evaluation will be conducted and what evaluation data needs and efforts are needed to support the evaluation. In addition, the Protocol provides background that administrators can use to determine when to intervene in the program design and implementation efforts to achieve continued and/or greater efficiency gains.

- Program Implementers – will use the Protocol to understand the evaluation that will be conducted on their programs and program components. Often, they will be required to provide data to support the evaluation. The Protocol will also provide background information that implementers can use to understand when to intervene to achieve continued and/or greater efficiency gains.

Key Inputs, and Outputs

There are several key Evaluation Protocol related inputs and outputs including the energy impacts caused by the program-induced changes. This section of the Protocol lists the key information inputs that are needed to conduct the evaluation and the key outputs that will be provided as a result of the evaluation.

Key Inputs

The major evaluation input metrics needed to conduct the evaluation efforts include:

1. Codes and Standards Program Theory and Logic Models,
2. Codes and Standards change descriptions,
3. Technology descriptions,
4. Program activity descriptions,
5. Identification of key codes and standards stakeholders,
6. Identification of the jurisdictions covered by the codes and standards changes,
7. Estimate of pre codes and standards technology adoption or penetration rates before changes to the code are made.

Key Outputs

The major outputs from the evaluation efforts include:

1. A listing of the technologies or practices influenced by the program that experienced an energy efficient code or standard change.
2. A listing of the code and standard changes that will be addressed in the evaluation. (Items 1 and 2 may be the same, but also may be different if the evaluation is addressing a subset of the changes.)
3. An estimate of the influence of the program on the code and standard changes for each technology or practice included in the evaluation.
4. An estimate of the naturally occurring market adoption rates for each technology or practice included in the evaluation.
5. An estimate of the date when each code or standard change would have occurred without the program for each technology or behavior included in the evaluation.
6. An estimate of the level of non-compliance expected for the technologies and practices covered in the evaluation over the period of time that savings are projected.
7. An estimate of gross and net market-level energy impacts for the program as a whole and for each technology and practice covered in the program and for each utility territory funding the program. This estimate of impacts should not exceed a 30-year effects lifetime.

Evaluation Methods

The evaluation of Codes and Standards programs requires an Evaluation Protocol that is guided by the Impact Evaluation Protocol. The primary approach to establishing an energy savings value for the Codes and Standards Program is to assess the energy impacts of the market adoption and decision changes caused by the code or standard change, and then adjust those savings to account for what would have occurred if the program never existed. The evaluation must identify the net energy impacts that can be directly attributed to the program's actions that would not have occurred over the course of the normal non-program influenced operations of the market.

The end result of the application of this Protocol is the identification of the net ex-post energy savings achieved from code and standard changes above and beyond what would naturally occur in the market through normal non-code/standard driven technology adoption behavior and through the normal cycle of codes and standards updating activities. The resulting net program-induced energy savings are the savings that are caused by the program's efforts.

The following sections of this Protocol describe the required efforts for evaluating these programs. We note that the evaluation of the Codes and Standards Program can be accomplished in a single multi-year study incorporating an assessment of the gross energy impacts from the code or standard changes, followed by the application of net adjustment approaches described in this Protocol to produce net effects. These two efforts can also be structured independently, as conducted in the 2005 study by the Hescong Mahone Group. That study relied on the energy impact estimates from previously conducted energy impact studies.

Evaluation Planning

Once an independent evaluator is hired, the evaluator must prepare a detailed evaluation plan that allocates the study's finite resources to maximize the value and use of the information collected. The plan must provide detailed task-level information and fully describe the data collection and analysis approaches that are to be conducted. The plan must be provided in enough detail that it can be replicated to achieve the same conclusions. As part of this plan, the evaluator must specifically address the various sources of relevant potential error and explain how the error will be mitigated⁷².

The Impact Evaluation Protocol will guide the gross market-level energy impact estimates, and the Sampling and Uncertainty Protocol will guide the gross market-level energy impact estimates and the approaches for identifying net adjustments to the gross savings. In conducting this evaluation, evaluators must assess, minimize, and mitigate potential bias and, when relevant, present the achieved level of precision (including relative precision, error bounds, coefficients of variation, standard deviations, and error ratios) for interpreting the data.

⁷² In the pre-1998 Protocols, there was no requirement to address these sources of error in the research plan. Evaluators only had to describe in the final report whether they had to address these various errors and, if so, what they did to mitigate their effects.

It is expected that a technology and behavior-specific code and standard application potential analysis will first be conducted to establish the population characteristics needed to inform the sampling plans associated with the evaluation. The potential analysis is an assessment that describes the current saturation and penetration of the specific technologies or behaviors that may be evaluated. The study then identifies the remaining potential that will be captured via a code or standard change. Alternatively, the evaluation will use market size estimates prepared by the Codes and Standards Program that have already projected the potential applications remaining in the market. If the program-developed potentials analysis is used, the evaluation contractor must first assess the methodological approach used by the program to determine the suitability for use in the evaluation effort and identify weaknesses in the projections that can influence the accuracy of the evaluation findings. If the assessment is found to be unreliable, the evaluation contractor will work with the Joint Staff to establish a methodology for estimating the market application potential and the characteristics of the markets needed to inform the study's sampling plans. This will help ensure that the gross and net energy impact estimates for the code or standard change is representative of the market in which the changes are to be measured.

The Joint Staff, and other outside stakeholders as deemed appropriate by the CPUC-ED, will review the evaluation plan submitted and discuss with the independent evaluator tradeoffs that are deemed necessary to maximize the reliability of the impact estimates. The Joint Staff can decide to modify the approach as necessary in order to increase precision or to improve the reliability of the study findings, or to have the plan meet budget or timeline considerations.

The evaluation plan will also identify any information that will need to be supplied by the utilities so that they will have advanced notice of what will be requested in an official data request once the study is launched.

Technology-Specific Code and Standard Change Theory

The first step in the evaluation process is to review the codes or standards change theories. The change theory is similar to a program theory, but it focuses on the measures included in the code or standard change, and the theoretical approach that the program is using to bring about the change. The change theory should present a story of how the program moves from the development of a change concept (for example, the need to change the code covering residential sidewall insulation in single family homes) to the completion of the code or standard change and a description of the savings expected. It should also include an estimate of the difference in the penetration of the code or standard-covered technologies between the pre-code adoption market and the post-code adoption market. The change theory should identify the activities that the program undertakes in its efforts to move from a change concept to a successful code or standard change. A code or standard change theory should be developed for each code or standard being changed. For example, if the code change focuses on duct sealing, there should be a duct sealing code change theory that describes the activities that will be used to bring about the change in duct sealing practice. The code and standard change theory should include:

1. A description of the technologies and measures affected by the change and the change being made.

2. A description of the program activities, efforts, and events associated with the change making process.
3. Identification of the key stakeholders the program needs to work with to influence the change, including their names, titles, organizations, addresses, phone numbers and where possible, their e-mail addresses⁷³. These should all be key market actors that are (or are expected to be) instrumental in bringing about or helping to bring about the change. These individuals should be grouped by their roles in the change making process (program management and implementation, code review and assessment, case study development, economic impact assessment, environmental impact assessment, market impact analysis, technology availability assessment, supply chain analysis, lobbying, policy review and development, skills analysis, etc.). By providing these examples, we are not suggesting that these classifications be used, but rather demonstrate that some form of responsibility classification be used so that the evaluation contractor understands their individual roles in the change process. Lists of individuals involved in the change efforts and their roles should be maintained throughout the program's implementation efforts and program managers should be ready to provide these lists to the evaluation contractor on request. These interviews will be conducted over the pre- and post-change period.
4. The outputs, products, efforts and activities from the program that are used to cause the change, identifying how they are used to affect or support the change.
5. The incremental and final outcomes from the program's change efforts and activities that have been or are planned to be accomplished.
6. The timelines associated with the program's change efforts, including the adoption dates of each change and the date the change is to apply. We expect that the program change timeline will be multi-year, because code or standard change efforts are launched at least two years before a formal adoption takes place, and at least three years will pass before they become effective in the market.
7. A description of the code and standard that has changed (after the official adoption), and an electronic or hard copy of the parts of the code or standard that are changed, with code or standard reference numbers to allow independent confirmation of the change.
8. A description of the jurisdictions covered and not covered by the code or standard changed, and any conditions that would exempt or prohibit a jurisdiction from implementing the code or standard. This should identify all the significant reasons why a code or standard may not be fully adopted within the jurisdictions affected by the code or standard.
9. A pre-code and standard change description of the penetration levels of the technologies covered in the code or standard in the targeted markets and a description of the expected penetration levels following adoption of the program-influenced code or standard. These penetrations should be provided for each of the markets being targeted for the code or standard.

⁷³ The names, addresses, and contact information of the people the program works with should be considered confidential information and protected from disclosure.

For assistance in understanding the nature of a program theory and the associated logic models, see the *California Evaluation Framework*, page 30. While the *Framework* does not detail what is included in a change theory, the codes and standards change theory should be similar to a program theory and the supporting logic model. However, the focus of the evaluation of the Codes and Standards Program is not at the program summary level, but instead is developed for each technology/measure targeted by the Codes and Standards Program. It is expected that the change theory for codes and standards programs will include theories on technologies/measures that are being successfully moved or have moved to a code or standard change. This condition allows the evaluation contractor to understand the full nature of the program operations and focus, including the approaches for technologies and measures that move from the concept stage to the code or standard change stage.

The code and standard change theory will be a key document used to guide the evaluation effort. Without the code and standard change theory, the evaluation contractor cannot fully understand the efforts, events, and key individuals that must be considered to develop the evaluation plan. This Protocol recognizes that the change theories will be developed and modified over time, as the program moves through the implementation process. The change theories developed early in the process are expected to be less specific and less “fleshed-out” than the theories developed mid-stream and during the final adoptions processes. The evaluation contractor will need to make sure that the change theories used to guide the evaluation efforts are the most recent theories. These Protocols require the program administrators to provide updated program change theories to the evaluation contractors immediately after they are developed or modified. However, the evaluation contractor should also confirm that they are planning the evaluation using the most up-to-date change theories.

Each code and standard change theory should be accompanied by a code and standard logic model that graphically displays each theory. The logic model will include the resources used by the program, the activities of the program, the outputs from the program activities, and the outcomes expected from the changed codes and standards.

These documents will be instrumental in estimating the level of influence of the program on the adoption of the specific codes and standards changes.

The evaluation contractor will request the program theory and logic models from the program administrator(s) immediately after the evaluation contract is negotiated. If the program staff has not developed the theory, the evaluation contractor will notify the program administrator(s) and the CPUC-ED that the code change theory is not available to guide the evaluation planning process and the evaluation planning efforts cannot proceed. At this time, the CPUC-ED will instruct the administrators of the program to develop the code and standard change theories and supportive logic models. The program administrator(s) will then develop the theories and the supporting logic models for the covered technologies.

If the program theories and logic models are not available at the time of the evaluation request, the administrator may elect to hire contractors to develop or help develop these materials. These materials must be delivered to the CPUC-ED within 40 days of the notice and be used to guide the development of the evaluation plan. The development of the evaluation plan should be

launched immediately after the evaluation contractor is hired, but not finalized until the program theories and logic models have been delivered and used to guide the evaluation planning efforts. Because program theories and logic models are “living documents” that change as program designs and objectives change, it is important that the most updated theories and models guide the evaluation plan. Alternatively, the CPUC-ED can instruct the evaluation contractor to work with the program managers to develop the change theories and the supporting logic models to guide the evaluation effort. If this step is taken, the Program Administrator must “sign-off” on the accuracy of the theories and the supporting models before they are used to guide the evaluation efforts.

Evaluation Approach

Identify the Evaluation-Covered Codes & Standards

In this effort the evaluation contractor, in coordination with the program administrators and Joint Staff, will identify the specific codes and standards that have been, in some way, influenced by the program’s activities, and identify those that will be incorporated into the evaluation effort. This assessment will use the code change theories, logic models and market actor information provided above, in addition to consultations with the program administrators and Joint Staff. Typically, the impact evaluation will focus on 5 to 25 changed portions of applicable codes and/or standards, depending on the number of code or standard changes that have been adopted, however, the actual number may be more or less than this range.

Not all energy-related code or standard changes are caused by or influenced by the Codes and Standards Program(s). These non-program changed codes or standards are not included in the impact evaluation. Similarly, not all codes and standards changes targeted by the program make it into a new code or standard, however the costs of these efforts should be included in the cost effectiveness evaluation of the codes and standards program, even if they have not yet become adopted by one or more jurisdictions.

The codes and standard changes that can be included in the impact evaluation plan and assessed in the evaluation are those for which:

1. The program has developed a code or standard change theory and supportive logic model,
2. The program-covered change has been adopted, or is expected to be adopted by at least one public jurisdiction (city, county, or state) who has made the code or standard a required or voluntary practice, and
3. The change theory provides a reasonable cause and effect relationship leading from a concept stage to an adopted code or standard, indicating that the program’s actions can be expected to have a positive influence on the adoption process. If there is disagreement on what constitutes “*a reasonable cause and effect relationship*,” Joint Staff will make the decision with advice from the program administrator and the evaluation contractor.

When these conditions exist, an assessment of the impacts of that technology or practice change will be included in the impact evaluation. However, the Joint Staff, after consulting with the program administrators and working in concert with the evaluation contractor, may elect to modify the code and standard changes addressed in the study as a result of expected or projected program actions.

Conduct a Codes and Standards Gross Market-Level Energy Impact Assessment

The evaluation contractor will conduct a load impact evaluation of the savings (kWh, kW, and therms of natural gas) expected from the technologies that are covered by the code and standard changes. This study is a gross market-level assessment that focuses on the total amount of savings that can be expected by the changes, regardless of the cause of those changes. However, this study only focuses on those changes that are targeted by the program and for which the code change theory explicitly identifies as being affected by the program's efforts.

In conducting this study the evaluation contractor will follow the Impact Evaluation Protocol to estimate savings from the technologies affected by the code or standard change. The "Basic Level of Rigor" for estimating gross energy impacts, as identified in the Impact Evaluation Protocol, is to be applied to assessing the gross market-level energy impacts. However, the Joint Staff can stipulate either more or less rigorous methods during the evaluation planning process if there is a need for more accurate savings estimates, if budget or timeline restraints requires a less rigorous approach, or if Protocol-covered evaluation findings that have already estimated the energy impacts for a given technology can be used to estimate market-level gross savings. The goal in establishing this requirement is to have flexibility in the evaluation design process to meet unforeseen barriers to the evaluation, but still establish a default level of rigor for which the estimates can be based. The evaluation contractor will work with the Joint Staff to set rigor levels consistent with the needs of the study, the study timeline and the evaluation resources.

The evaluation contractor may not need to conduct an impact evaluation assessment on a particular technology or practice if that technology or practice has already been evaluated using a reliable impact assessment approach similar to the approaches covered in the Impact Evaluation Protocol (2006). When previous evaluation findings can be directly used or modeled (simulated) to reflect the use and application conditions associated with the changed codes and standards, that approach should be used if it results in a reliable energy savings estimate. Likewise, the evaluation contractor may not need to conduct an impact evaluation on a particular technology or practice if a review of the program's estimates of energy savings, and the supporting documentation and case studies, are found to be reliable. In this case the evaluation contractor should review the program's estimated savings and, in consultation with Joint Staff and the program administrators, discuss the threats to validity associated with the estimation approach and determine if the approach is reliable enough that the evaluation contractor can use the estimates, or if they can be made more reliable through additional engineering adjustments, modeling or modeling changes, additional field M&V, or application testing efforts. The purpose of allowing the use of previous evaluation results and of the program's energy saving estimates is to not expend evaluation resources if reliable energy savings projections can be constructed by using previous work.

If there are no previous impact evaluation studies associated with a specific code or standard change that can be used, or adjusted and used, and if the program's energy savings estimates are found to be unreliable or have significant threats to validity making them unreliable even with addition modeling, M&V efforts, or field testing, the evaluation contractor is to develop a plan to assess the energy savings for that technology or practice using the Impact Evaluation Protocols to develop the evaluation approach.

It is expected that as the Energy Impact Protocol (2006) is adopted and used, more and more technologies will have been evaluated under the Protocols in which the results can be used or adjusted to reflect expected code and standard application conditions, thereby reducing the need for new technology evaluations to feed the codes and standards gross market-level impact estimates.

As noted earlier, the default approach for conducting the market-level energy impact assessment is set at the Basic Level of Rigor as specified in the Impact Evaluation Protocol for estimating gross program impacts unless the Joint Staff or the CPUC-ED has assigned a different level of rigor for a given technology. In making the rigor assignments, the Joint Staff will consider past evaluations and their energy savings estimates for covered technologies and the potential to use these study results, the need for different levels of accuracy in the market-level energy assessment for individual technologies, the available budget to support the assessment and the timeline for the evaluation, in addition to other criteria. These requirements mean that at a minimum:

1. Simple engineering model estimation approaches, or
2. Normalized annual consumption approaches will be used, unless
3. The CPUC-ED or Joint Staff have approved an alternative approach based on the criteria discussed above.

The results of this assessment will be an annual energy savings estimate covering the first year of code or standard adoption for each technology or behavior change covered in the change theory. This estimate will be based on the expected penetration rate associated with each change across the market sectors for which the code or standard change applies, assuming that it would impact all installations covered by the change. In assessing the savings it will be necessary for the evaluation contractor to estimate the increase in adoption of each technology or behavior change resulting from the code or standard change. This assessment will most likely involve the use of projected construction levels grounded on historic construction patterns, estimated retrofits and change-outs driven by normal market forces, and other estimates of change for each of the changes. This annual savings will then be projected into the future to construct a time-sensitive estimate of gross savings.

In assessing the gross market-level energy savings it will be important for the evaluation contractor to understand that the code or standard changes supported by the program's efforts may not be consistent with the newly adopted changes. That is, the program may focus its efforts on a more aggressive or less aggressive energy efficient change to the code or standard than what is actually adopted. As a result, the gross energy savings assessment must focus on the changes made to the adopted codes and standards that were influenced by the program, rather than the changes recommended by the program. Likewise, the evaluation contractor must check

to see that the program-influenced changes are still in force. Just as codes can change to be more energy efficient, they can also change to be less energy efficient.

In assessing the gross market savings the evaluation process should disaggregate the savings assessment efforts into the specific installation, construction or purchase changes being made as a result of the code or standard change. This may mean disaggregating the savings analysis into measure groups or small clusters of measure groups rather than aggregating multiple measures and practices into large groups⁷⁴.

Once the gross market-level energy impacts are identified, the following approach will be applied to develop an estimate of net program effects.

Estimate the Program's Influence on the Adoption of Codes & Standards

Once the gross market-level energy savings estimates are established, they must be adjusted to account for the influence of the Codes and Standards Program on the code or standard change. The program may be only minimally responsible for a given change, or may have had a significant influence on the code and standard adoption process⁷⁵. For each technology or behavior, the evaluation contractor must establish a percent attribution factor for the savings that can be attributed to the program. These percentages can range from no influence (0% if the program had no tangible influence on the change) to a significant influence potentially approaching 100 percent (if the program was the primary influencing factor driving the change).

A stakeholder interview-based preponderance of the evidence approach will be used for this process. This process will identify key stakeholders and conduct multiple interviews with these stakeholders at different points in time along the adoption path, during both the pre-adoption and post-adoption period.

The evaluation contractor will conduct interviews with a representative sample of the key stakeholders identified earlier (see item 3 in the Technology-Specific Code and Standard Change Theory above) and use the results of these interviews, along with reviews of program materials and documents (including lobbying documents, staff reports, case studies, and staff and stakeholder correspondence as available) and attendance at program meetings and key events associated with the adoption process (to the extent possible and practical) to assign causation percentages for the change to various change agents identified by the stakeholders, including direct or indirect efforts of the program. In making these attribution assignments, the evaluation contractor will want to consider the potential bias of the individuals interviewed and of the information reviewed, and cross-check stated opinions with applicable documents and the opinion of other stakeholders, in order to test the causal relationships between actions and results. The evaluation contractor should make as objective an assignment as possible. The evaluation contractor will assign weights to the opinions of the stakeholders based on a review of all available information (noted above). The contractor will assign higher weights to those who

⁷⁴ Note: a previous study aggregated the assessment into one change assessment cluster that represented 66% of the savings even though the change represented different measures, approaches and technologies. The study should disaggregate the assessment to the extent possible and practical given the evaluation needs and resources.

⁷⁵ The assignment of attribution of cause is to assess energy savings via the evaluation approach. It is not placed in this Protocol to establish the program's NTG values or to change the ex ante projected savings.

are most likely to have a complete understanding of the change efforts and processes relative to a specific technology or set of technologies and who are more likely able to accurately judge the relative causes of the adoption of the new codes or standards. This will allow the attribution of change to be more informed by those who are in a position to best judge the reasons for the change. Utility and other program staff and contractors hired by the program should be included in the sampling approach and be interviewed. As with these and other individuals interviewed, the evaluation contractor will keep in mind the potential biases that may be associated with any single individual. In the weighting process, significant weights should be applied to the opinions of non-program stakeholders who are instrumental in the statewide jurisdictional decision processes to adopt a code or standard change and to advisors or key stakeholders informing this process. In selecting a sample of interviewees, the Sampling and Uncertainty Protocol should be followed, with the sampling method determined at the individual code or standard change level. The interview process should be structured to conduct both pre-change and post-change interviews.

The interview protocol and the interview guide should be designed to be objective and rely on the opinions of the key stakeholders. The interview guide should be a prompted guide, so that the interviewee is not placed in the position of trying to identify all the different causes for the change. The evaluation contractor will develop a list of program and non-program associated change agents/causes based on a review of the change theories and interviews with a small but adequate sample of evaluation contractor-selected program and non-program stakeholders. The sample selection for these interviews does not have to follow the Sampling and Uncertainty Protocol.

The evaluation contractor will plan the sample selection for the stakeholder interviews to focus on the program-identified stakeholders contained in the change theory documents or other associated documents. However, the evaluation contractor will use a “snowball” sampling approach in which the sampled interviewees will be asked to identify additions to the sample of individuals the interviewee indicates were instrumental in the change consideration or decision process. The evaluation contractor will target an additional 20 percent of the interview sample points to interviewing stakeholders recommended by the interviewees who are not on the change theory stakeholder list. If the evaluation contractor is unable to obtain an additional 20 percent, the contractor will conduct as many of the additional interviews as possible and state in the evaluation report that they were unable to identify or interview an additional 20 percent.

The results from the interviews will be aggregated and used to assign technology and behavior change attribution of the changes caused by the Codes and Standards program. The results of this process will be a percentage distribution of the causes for each change across the stakeholder-identified reasons for the success of the newly adopted code or standard change for each of the technologies or behaviors covered.

It is expected that there will be significant levels of interview overlap across the technologies and behaviors so that a single interview may cover several technologies or behaviors related to a code change or changes. This sampling process assures that adequate samples will be selected for each technology or behavior-associated change and that the attribution will be based on program-identified and stakeholder-identified change agents.

Once the attributions have been established at the technology level, the evaluation contractor will multiply the energy savings for each technology or behavior by the attribution score to identify the gross market-level energy impacts that were caused by the Codes and Standards Program. This savings estimate will be further adjusted to account for net program effects (see below).

The timing of the estimation of the program's influence is critical to the success of the evaluation. The attribution assessment must be started very early in the Codes and Standards Program cycle, but not completed until the adoption process has been completed for the changes being evaluated. The technology or behavior change selections and the associated code and standard development efforts for the 2008 codes and standards began in the fall of 2005 and will continue through early 2006. In order for the attribution efforts to be based on recent knowledge, the interviews must be conducted during the technology selection and demonstration development process (as appropriate) and again when the adoption process is complete. This means that the attribution assessment may need to be launched years before the program experiences its first code or standard associated savings⁷⁶.

In assessing the program's influence on the adoption process, the evaluation contractor should consider a number of program and market conditions and activities that influence the adoption process and the associated adoption decisions relative to the individual changes. In considering these changes the Protocols references the Codes and Standards white paper⁷⁷ in which different adoption influence weights were used to assign attribution. While this white paper should be examined in the evaluation planning process, the evaluation contractor should be careful not to select program or market condition weighting criteria that correlates with or overlaps among the weighting metrics so that the weighting approach acts to double-count adoption influence across more than one of the weighting criteria.

Estimate Net Program Induced Energy Impact

The gross market-level energy impacts that were caused by the Codes and Standards Program must be adjusted to account for naturally occurring market adoption changes, normally occurring codes and standards revisions, and non-compliance with the new codes and standards. These adjustments are discussed below and need to be made in the order prescribed in this Protocol.

Naturally-Occurring Market Adoption

The first adjustment to the gross energy savings estimate identified above is an adjustment to account for the naturally occurring market adoption rates. New energy efficient products are likely to penetrate and be adopted by at least a portion of the market even without the Codes and Standards Program. As a result, the projected naturally occurring adoption and penetration, which would occur without the program, needs to be subtracted from the program's gross energy impacts.

⁷⁶ This means that evaluations of codes and standards programs conducted in the first years following the issuance of this Protocol will be operating in a "catch-up" mode because the program will have already launched the change efforts on which the first evaluation will focus.

⁷⁷ Codes and Standards Program Savings Estimate, August 1, 2005 (or most recent revision), Heschong Mahone Group, page 8. CALMAC SCE0241.01.

Naturally occurring adoption rates for premium energy efficient products typically occur in an “S” shape pattern that never reaches 100 percent penetration as long as there are alternative technologies in the market. This is especially true when the alternatives are lower cost technologies. Some energy efficient technologies may never capture a majority of the market share without a mandatory code or standard. Others may move to capture the majority of the market without a code or standard. However, there is likely to always be some level of increased penetration of a superior product that delivers benefits to a user, up to a point of product demand saturation, based on the characteristics of the product and the alternative choices in the market. Similarly, some customers never adopt a new product regardless of the benefits of the product. These customers are typically labeled as “laggards” within the technology adoption literature.

This step requires the evaluation contractor to establish expected adoption curves for each technology included in the impact assessment. The evaluation contractor will use a range of approaches to establish the estimated penetration curves, including conducting literature searches on the penetration rates of similar technologies with similar product characteristics, the use of expert opinions on the expected penetration rates in the absence of a requirement to use the technology, relevant market data and other approaches as deemed appropriate in the evaluation planning effort.

The evaluation contractor will then adjust the projected savings to account for the naturally occurring adoption for each technology covered in the assessment.

Non-Compliance Adjustment

The second adjustment to gross savings is an adjustment for non-compliance. Since not all buildings or appliance decision makers will fully comply with the newly adopted codes or standards, these lost savings must be subtracted from the gross estimate.

In the real world, there is often a range of appliances or measures present in the market, some falling below the standard and some above the standard in their energy efficiency levels. Similarly, technologies that do not comply with the new code or standard are often stocked and sold in the market regardless of the requirements adopted. For example, while programmable thermostats are now required in California for most space heating and cooling applications, it is easy to acquire and install non-compliant thermostats because of the stocking and sales patterns of a wide variety of wholesale and retail outlets, including internet sales. In some cases, if permits are not required or obtained, the codes and standards enforcement mechanisms associated with the building inspection process may not be applied, enabling non-approved installations to occur. Likewise, it is difficult to inspect code-covered applications of measures such as insulation once the construction is completed to enforce code compliance, making this measure difficult to inspect and enforce.

In order to comply with the Evaluation Protocol, the evaluation contractor must estimate non-compliance across the technologies being assessed and adjust the anticipated savings for the net non-compliance rate over time. For technologies that do not comply, but are easily available in the market, the non-compliance rate may be high. However, for other technologies that are typically inspected as part of the construction or retrofit process, the non-compliance rate may be low.

To establish the rate of non-compliance the evaluation contractor will conduct interviews with a set of building architects, engineers, contractors, product wholesalers and retailers and installation contractors. The evaluation contractor will design a sample plan consistent with the Sampling and Uncertainty Protocol to match the technologies being assessed. Because compliance is measure-specific, samples will be set at the technology level within each code or standard changed. In developing this adjustment the contractor will need to be sensitive to differences in compliance rates across the state and over time. As a result, the evaluation contractor should consider approaches for adjusting for local differences in compliance rates, such as establishing and using compliance assessment jurisdictions. These approaches will be coordinated with and approved by Joint Staff before they are implemented.

The evaluation contractor will also assess the availability of non-compliant technologies in the market by examining the stocking practices of selected suppliers of the technologies. For example, if a building products supplier stocks 30 percent non-compliant technologies, the non-compliance rate for that technology can be assumed to be 30 percent for their customer market, unless there is evidence to the contrary collected during the interview efforts. The evaluation contractor will suggest ways to conduct the stocking assessment and can include such approaches as visits to suppliers to examine the stocking mix or interviews with suppliers to estimate their stocking mix.

The evaluation contractor will then assess the results of the interviews, the examinations and other assessment approaches suggested by the evaluation contractor and approved by the Joint Staff and estimate the rate of compliance for each technology or behavior change. The estimate will not be a single fixed level, but will be time-adjusted, so that the expected rate of non-compliance will change over time. To arrive at the time-adjusted compliance estimate the evaluation contractor should rely on projections provided by the interviewees.

It is important for the evaluation contractor to focus on identifying net compliance adjustments during this assessment and take into account the pre-change compliance rate for a given change condition. There may be substantial portions of the market that are not in compliance before the change and are not in compliance after the change. Likewise, a non-compliant rate before the program may have the same non-compliant rate after the change. The evaluation contractor is expected to develop plans that provide for net compliance changes over time. The contractor will coordinate with Joint Staff in this effort.

The evaluation contractor will then adjust the projected savings to account for the estimated levels of non-compliance.

Normally-Occurring Standards Adoption

Next an adjustment to the gross savings needs to account for the normally occurring codes and standards change process. A primary effect of the Codes and Standards Program is to accelerate the time it takes for the CEC and other jurisdictional organizations to update current codes and standards or adopt new codes or standards. The CEC employs a three-year update cycle, keeping the standards up-to-date and cost-effective as market conditions change. However, without the Codes and Standards Program resources, the updates might not encompass the same type of technology analysis and change considerations. It is reasonable to assume, therefore, that the

standards adopted by the CEC or other jurisdictions would have been adopted in the normal course of events, but over a much longer period of time. The energy savings from the Codes and Standards Program should only include the savings from the codes and standards implemented as a result of the program's efforts for the period of time that they would not be covered by a revised code or standard during the normal course of the update cycle.

In order to establish the estimated time at which the CEC and other jurisdictions would have adopted or created a code or standard without the program, the evaluation contractor must establish a panel of experts who are familiar with and involved in the code change efforts. This panel will consist of CEC program staff, CEC code and standard update staff, code and standard public officials within other jurisdictions, and other experts as deemed appropriate by the evaluation contractor and approved by the CPUC-ED or the Joint Staff. The evaluation contractor will then conduct a minimum two-round Delphi⁷⁸ assessment with this expert panel to arrive at a projected date that the CEC would be expected to implement a new code or standard in the absence of program initiatives. This process should cover each technology or behavior in the assessment. It is expected that the size of this panel will be between 10 and 20 experts.

Once the estimated timeline for each code or standard change is established, the energy savings for the technologies and behaviors changed as a result of the code and standard changes will not be counted beyond that projected date, but in no event will the savings be counted beyond a 30-year period. This step sets an end-date for the period of time that savings can be counted for each code or standard change.

Actual Construction and Retrofit True-Up

The energy savings estimates produced from this Protocol are based on a single assessment of the gross energy savings for a single year projected into the future. However, not all years are the same. The economy and other changes (interest rates, unemployment, consumer confidence, etc.) affect the rate at which technologies are adopted and used, and thereby influence energy savings. As a result, it is necessary that the CPUC-ED may elect to periodically issue a new RFP to conduct an update of the projected savings to account for actual savings. When the CPUC-ED requests an update, the evaluation contractor will assess the market and update the savings projections to account for actual construction and adoption.

It is not possible to accurately estimate savings without knowing how much construction was actually accomplished following a code or standard change. There are several ways to adjust the energy savings projections to account for actual construction and a preferred approach is not specified in this Protocol. However, a true-up of actual construction is needed to help increase the accuracy of the savings estimate over time. If the true-up evaluation is conducted in the 5th year following the code change, then the true-up should contain estimates of actual construction for the first 4 years of which permitting and building records could be assessed. Once the evaluation has a history of actual construction, the new projection of future construction (to estimate future savings) can be based on the historical construction. Once the projection is

⁷⁸ Delphi assessment is an iterative process that involves repeated rounds of information gathering across a selected group of experts. Responses to one round are summarized and developed to feed the next round of information gathering. The purpose of the Delphi is to seek agreement across the group of experts.

established, the savings can be projected and the adjustments can then be subtracted (or added, depending on actual construction data) from these original projections to obtain net realized past savings and the updated projected future savings based on the updated estimate. When an update is requested the CPUC-ED or the Joint Staff will work with the evaluation contractor to identify an approach to be used. This approach may be based on construction industry statistics (e.g., annual real estate construction estimates), building construction databases (e.g., the Dodge database and/or Construction Industry Research Board (CIRB) reviews of building permits for a set of representative jurisdictions), assessments of sales data if the data can be reliably obtained (a historic problem for sales data collection) or other approaches.

Multiple-Counting of Energy Savings Adjustment

To make sure that the savings from code change covered measures, practices and purchases are not counted more than once, no energy efficiency or demand management/response programs that offers code or standard change covered measures, equipment or practices is permitted to count the savings from these measures, practices or purchases toward their energy savings goals once the codes and standards evaluation has documented the savings from these efforts, unless those savings are from Code Compliance Enhancement programs. The Code Compliance Enhancement Program evaluation will then document savings beyond what is achieved as a result of the code and standards change⁷⁹.

Measure Life Adjustments

This Protocol excludes an adjustment for measure life. It is assumed that once a measure is adopted as a result of a code or standard change, the behavior will be repeated until that code or standard is eliminated or updated. However, even if the code change is updated, the savings from the measures are still provided. Likewise, new evaluations will document the increased efficiency of the updated codes or standards. In addition, the inclusion of normal market adoption rate adjustments and normal code and standard change revisions will act to significantly reduce the savings over time. However, the energy savings provided via the use of this Protocol shall not be projected beyond 30 years.

Impacts by Utility Service Territory

Once the statewide estimates of adjusted net savings have been estimated, an allocation of the savings to the utility service territories can be made. This assignment of savings will be based on assigning savings to a utility for measures that are actually installed within their service territory. The allocation will be based on the distribution of new home construction, nonresidential construction square footage, and appliance sales forecasts within each service territory such that the total savings across the territories equals 100% of the adjusted net savings estimated from the program less the savings from the local jurisdictions that had implemented or were in the process of implementing the adoption of changes covered in the scope of the evaluation. This Protocol condition means that the evaluation contractor will need to conduct a survey of at minimum the 20 most populated (or preferably and provided the data is available the 20 jurisdictions with the highest numbers of building-starts) local jurisdictions within each IOU service territory to assess if that jurisdiction was substantially in the process of converting their code or standards to the

⁷⁹ See Protocol steps for assessing Compliance Enhancement Programs located at the end of this Protocol.

covered changes or had completed this effort at the time the program was advocating for the change. The evaluation contractor must obtain approval from the Joint Staff on the jurisdictions to be surveyed for each code or standard change. It is expected that there will be substantial overlap among the identified jurisdictions and that most of the targeted jurisdictions will be surveyed for more than one of the program's covered changes. In selecting the local jurisdictions to survey, the evaluation contractor will survey enough jurisdictions to be able to reliably measure the program's net effects.

Because the construction, retrofit and sales markets change over time, this assessment and the adjustment approach will need to be trued-up periodically. These refinements will be specified by the Joint Staff, or the CPUC, in order to allocate savings over time based on market conditions.

Reporting

The evaluation report will be provided in compliance with the Reporting Protocol. A draft report will be provided for review and comment to the stakeholders (see Reporting Protocol). Once comments are provided on the draft report the evaluation contractor will work with the Joint Staff to finalize the report. Once the final report is accepted by the Joint Staff, the evaluation contractor will construct an abstract consistent the instructions contained on the CALMAC.org web site and post the report.

Summary

This Protocol describes a way to estimate the gross and net energy impacts from the Codes and Standards Program. It begins with the review of the program change theory and logic models and the development of an evaluation plan. The implementation of the plan consists of estimates of gross market-level impacts for each technology and behavior adjusted to account for naturally occurring market changes, non-program induced code and standard revisions, and code compliance rates for each technology. The evaluation delivers net impacts for each technology and for the program as a whole, and then distributes the energy impacts to the participating utility companies.

This Protocol is prescriptive in nature, but allows for the use of new techniques or approaches when approved by the CPUC-ED or Joint Staff. As a result, it does not impede the evolution of evaluation approaches.

The Codes and Standards Program Evaluation Protocol is guided by the Impact Evaluation Protocol, the Sampling and Uncertainty Protocol and the Reporting Protocol.

The following table provides a summary of the Protocol that can be used to guide the evaluation efforts once the detailed contents of the Protocol are well understood.

Summary of Protocol-Driven Codes and Standards Evaluation Activities

1	Joint staff selects an evaluation contractor to implement the Codes and Standards Program evaluation.
2	The evaluation contractor reviews the program change theories and the program logic models, identifies the technologies or behaviors that can be evaluated via the Protocol, constructs a draft evaluation plan and submits the plan for approval to the CPUC-ED. The contractor works with the CPUC-ED on the development of the draft evaluation plan and rigor levels. The plan must use the Impact Evaluation Protocol, the Sampling and Uncertainty Protocol and the Reporting Protocol in the development of the evaluation plan and in the implementation and reporting efforts.
3	The CPUC-ED works with the evaluation contractor to finalize and approve an evaluation plan from which the contractor can begin the evaluation effort.
4	The contractor conducts an assessment of the gross market-level energy impacts for each code and standard covered technology or behavior being evaluated consistent with the rigor level assignments.
5	The contractor determines the influence of the program on the adoption of each code and standard covered in the study and allocates adoption attribution. The assessment uses an interview approach for this assessment. This assessment is accomplished as early in the code change cycle as possible but preferably in the technology selection and demonstration phase of the cycle.
6	The contractor estimates naturally occurring code and standard covered technology or behavior adoption rates based on literature reviews and interviews with experts.
7	The contractor adjusts the gross market level energy savings estimates to account for the net adjustment factors for naturally occurring technology adoption, naturally occurring code change, and non-compliance. This approach nets out the influence of non-program-induced impacts from the gross market-level impacts for each technology.
8	The contractor estimates the timeline associated with adoption of a code and standard without the program, using a Delphi approach with an expert panel.
9	The program administrators remove savings estimates from their programs for code-covered measures.
10	The evaluation contractor assesses the construction and sales efforts for each utility company service territory and allocates savings by IOU based on the construction and sales estimates.
11	The contractor reports the results of the evaluation to the CPUC-ED and Joint Staff consistent with the provisions in the Reporting Protocol.
12	Once the report is accepted by the CPUC-ED, the contractor develops abstracts and posts the report on the CALMAC web site following the CALMAC posting instructions.
13	As needed, the CPUC-ED or the Joint Staff can request the evaluation contractor to update and report the actual energy savings over time consistent with the Protocol. Updates can be conducted with a different evaluation contractor than those doing the

	original assessment.

Note: the steps included in this evaluation summary table must be accomplished in accordance with all the requirements within the Codes and Standards Protocol in order to be in compliance.

Code Compliance Enhancement Programs

To conduct energy impact evaluations of programs designed to influence the rate of compliance of code-covered measures the evaluation contractor should not follow the Codes and Standards Program Evaluation Protocol presented above, but should follow this Protocol specifically designed to estimate the energy savings from these programs.

Because the California IOU portfolios have not included code compliance improvement programs in their portfolios, and because these programs have yet to be evaluated to the extent that a standard evaluation approach can be reliably identified, this Protocol allows a wide variety of methods and approaches for assessing the savings from these efforts. After several of these evaluations have been conducted and the success of the approaches documented, a standard approach may be developed and added to the current Protocols. Until that time the following guidance will be used to structure and implement the evaluation of code compliance enhancement programs.

Definition of a Code Compliance Enhancement Program

A Code Compliance Enhancement Program (CEP) is any energy efficiency, demand reduction or demand management program whose primary purpose is to increase the level of customers complying with a code requirement that saves energy (kWh, kW, therms).

What this Protocol is Designed To Do

This Protocol establishes a framework under which CEP programs are to be evaluated to assess energy impacts. This Protocol does not establish program designs, program design criteria or program development approaches.

Code compliance enhancement programs can be incentive programs that are designed to increase compliance by providing incentives to customers to do what is required, educational programs to make customers or trade allies aware of the code and the need for compliance, training programs to train customers or trade allies how to comply, enforcement programs that take enforcement actions against non-complying property owners, or other types of program designs. These programs may also involve more than one type of delivery strategy.

Joint Staff Responsibilities

Because CEP are not (at this time) part of the suite of energy program services delivered in California, the Joint Staff are responsible for determining when to evaluate a CEP program and how that evaluation should be conducted and reported. However, that evaluation must employ the following approaches unless other approaches are requested and approved by the Joint Staff.

Draft Evaluation Plan

The evaluation contractor will prepare a draft detailed evaluation plan and submit that plan for review and approval to the Joint Staff. The plan should provide for a time-series measurement approach that can be replicated at different times over the implementation period. Joint Staff will review and comment on the evaluation plan and will work with the evaluation contractor to focus the plan on the evaluation objectives of the CPUC. This plan will serve as the approval process for launching the detailed evaluation planning efforts.

Program Theory Review and Assessment

The evaluation contractor will review and assess the program theory provided by the program administrator. This review will be focused on understanding the approach the program is taking to effect a compliance change and the activities that are employed to accomplish the program's objectives. Once the program theory has been assessed the evaluation contractor will modify the draft evaluation plan and submit the plan to Joint Staff for review and approval. One purpose of the program theory review is to allow for the examination of the program theory to feed the evaluation planning process so that the evaluation contractor can identify key measurement points on which the program needs to focus. Once the draft evaluation plan has been updated from the program theory review effort and approved by the Joint Staff the evaluation efforts can be launched.

Pre-Program Compliance Rate

The evaluation contractor will work with the Joint Staff to develop an approach for measuring the pre-program compliance rate for the measures covered by the program. This approach should focus on assessing the condition of the market and taking measurements that allow the evaluation contractor to identify the level of pre-program non-compliance within the geographical areas targeted by the program. The Codes and Standards Evaluation Protocol incorporates instructions on assessing compliance rates for evaluating Codes and Standards programs. These instructions are incorporated into this Protocol as a guidance resource for identifying non-compliance rates during the pre-program period. The Joint Staff and the evaluation contractor are free to develop other methods if, in the opinion of the Joint Staff, the alternative approach can be expected to be more or equally reliable to the approach presented in the Codes and Standards Protocol.

The purpose of this activity is to establish the baseline from which post-program changes in construction practice or measured installed can be assessed. It is expected that this assessment will need to be sensitive to local jurisdictions and changes in compliance within the local jurisdictions. The outcome of this effort will be the identification of the level of compliance for each program-targeted code change within the market sectors and jurisdictions on which the program's efforts are focused.

It is expected that the pre-program compliance rates will be set at some level of detail that will allow the evaluation to identify jurisdictional differences in compliance rates.

Post-Program Compliance Rate

At a period of time to be determined by the Joint Staff the evaluation contractor will again apply the same strategy used to assess pre-program compliance rates within the geographical areas targeted by the programs. These areas on which the evaluation will focus must also be the same areas of the state that the pre-program compliance assessment was focused so that the jurisdictions examined in the pre-program assessment match the jurisdictions examined in the post-program assessment. The primary purpose of this effort is to document the compliance rate after the program has been implemented long enough for an expected change in compliance to be measurable.

The time periods for the post-program compliance rate assessment will be set periodically over the program implementation period to allow results to be tracked over time and reported consistent with the reporting needs of the CPUC. For some measures and programs this may mean an assessment every six months, for others the assessment can be done annually, while for others the assessment may be needed every few years. The evaluation contractor in coordination with the Joint Staff will identify the periods in which the post-program compliance should be assessed.

Adjustment For Naturally Occurring Compliance Change

The natural compliance rate for most code requirements will change over time. Normally, compliance is expected to be lower on the date the change first applies. This is then followed by a period in which compliance rates increase and begin to stabilize as the change is structured into the market and local code officials and trade allies change their approaches to comply with the new code. Because the compliance rates change as a normal course within the market operations, the normal compliance rate that would have occurred without the program must be adjusted out of the calculation for net program compliance changes. The evaluation contractor will work with the Joint Staff to identify an approach for identifying normal compliance change rates. Because these programs have not been implemented or evaluated in the past, a prescribed approach for identifying the rate of naturally occurring compliance is excluded from this Protocol. Joint Staff may wish to employ trade ally surveys, expert panels, code official interviews, measure sales tracking approaches or comparison areas where the program services are not offered. Each of these approaches has their own strengths and weaknesses that should be assessed and considered in the planning process. Joint Staff or the evaluation contractor may wish to suggest other approaches for consideration. However, the Joint Staff must approve the procedures for identifying naturally occurring compliance change before the effort is launched.

Net Program-Induced Compliance Change

Once the pre-program and at least one round of post-program compliance assessments is conducted, the evaluation contractor will assess the net change in compliance across the jurisdictions targeted by the program. The evaluation contractor will work with the Joint Staff to identify the approach to be taken in this assessment, however it is expected that the approach will be a simple jurisdiction controlled net change assessment.

Assessment of Energy Savings

Once the net assessment of change is identified, the evaluation contractor will use the savings estimates provided from the codes and standards program evaluation for the same code covered measures included in the CEP. If the codes and standards evaluation effort is not completed in time for the CEP evaluation, the Joint Staff will decide to delay the completion of the CEP evaluation or launch the impact assessment approach prescribed in the Codes and Standards Evaluation Protocol for assessing net energy impacts from the code covered changes. However the Joint Staff can consider the use of other approaches, such as the assignment of DEER estimated savings for the covered measures, or an engineering-based assessment to estimate the probable energy savings, if these approaches are considered reliable predictors of the savings associated with a specific change. Other approaches can be applied at the request or approval of the Joint Staff, but these methods should focus on obtaining reliable savings estimates consistent with the available evaluation budget and the study timelines.

The net assessment procedures used must take into account the measures that would have been installed or constructed without the CEP program and the energy consumption difference between what would have been installed or constructed compared to the code-required efficiency levels.

Recommendations for Program Changes

The evaluation contractor is also to provide recommendations for program changes that can be developed as a result of the examination of the program theories, and the implementation of the evaluation assessment efforts. While this evaluation is not a process evaluation, the evaluation contractor may be able to provide valuable change recommendations that can be considered by the program administrator.

Cost Effectiveness Assessment

The evaluation contractor will conduct a cost effectiveness assessment using the program cost data reported to the CPUC in the monthly or quarterly program progress tracking cost reports submitted by the administrator to the CPUC. The evaluation contractor will conduct a TRC and a PAC test consistent with the approach provided in the Standard Practice Manual. The results will be reported in the evaluation report.

Reporting of Evaluation Results

The evaluation report should follow the Evaluation Reporting Protocol to meet the timelines and deliverable dates specified in the approved evaluation plan. The deliverable dates will take into consideration the reporting needs of the CPUC across the multi-year program implementation period.

References

- Belzer, D., K. Cort, D. Winiarski, E. Richman, and M. Friedrich, (2002) *Analysis of Potential Benefits and Costs of Adopting ASHRAE Standard 90.1-1999 as a Building Energy Code in Illinois Jurisdictions*. Report prepared for U.S. Department of Energy by Pacific Northwest National Laboratory.
- Cort, K. and D. Belzer, (2002) *Statewide Savings Projections from the Adoption of a Commercial Building Energy Code in Illinois*. Report prepared for U.S. Department of Energy by Pacific Northwest National Laboratory.
- Cort, K., D. Belzer, D. Winiarski, and E. Richman (2002) *Analysis of Potential Benefits and Costs of Updating the Commercial Building Energy Code in Iowa*. Report prepared for U.S. Department of Energy by Pacific Northwest National Laboratory
- Cort, K., D. Belzer, M. Halverson, E. Richman, and D. Winiarski (2002) *Analysis of Potential Benefits and Costs of Adopting ASHRAE Standard 90.1-1999 as a Commercial Building Energy Code in Michigan*. Report prepared for U.S. Department of Energy by Pacific Northwest National Laboratory.
- Jacobs, P., and D. Roberts. (2001) "Estimated Impacts of the Non-Residential and Appliance Standards" in *CA IOU Codes and Standards Earnings Claim Framework Final Report*. San Francisco, CA: Pacific Gas and Electric Company.
- Heschong Mahone Group. (2005) *Codes and Standards White Paper on Methods for Estimating Savings*, prepared for Southern California Edison Company.
- Heschong Mahone Groups. (2005) *Codes and Standards Program Savings Estimate*, Prepared for Southern California Edison Company.
- Heschong Mahone Group. (2001) *CA IOU Codes and Standards Earnings Claims Framework, Final Report*. Prepared for Pacific Gas and Electric Company.
- Lucas, R. (2001) *Assessment of Impacts from Updating Idaho's Residential Energy Code to Comply with the 2000 International Energy Conservation Code*. Report prepared for U.S. Department of Energy by Pacific Northwest National Laboratory.
- Lucas, R. (2001) *Assessment of Impacts from Updating Kentucky's Residential Energy Code to Comply with the 2000 International Energy Conservation Code*. Report prepared for U.S. Department of Energy by Pacific Northwest National Laboratory.
- Lucas, R. (2002) *Assessment of Impacts from Adopting the 2000 International Energy Conservation Code for Residential Buildings in Illinois (Draft)*. Report prepared for U.S. Department of Energy by Pacific Northwest National Laboratory.
- Lucas, R. (2002) *Assessment of Impacts from Updating New Mexico's Residential Energy Code to Comply with the 2000 International Energy Conservation Code*. Report prepared for U.S. Department of Energy by Pacific Northwest National Laboratory.
- Nittler, K., Wilcox, B. (2001) "Estimated Impacts of the Residential and Appliance Standards" in *CA IOU Codes and Standards Earnings Claim Framework Final Report*. San Francisco, CA: Pacific Gas and Electric Company.

Effective Useful Life Evaluation Protocol (Retention and Degradation)

Introduction

One of the most important evaluation issues is how long energy savings are expected to last (persist) once an energy efficiency measure has been installed. The Effective Useful Life (EUL) Evaluation Protocol was developed to address this issue and should be used to establish the period of time over which energy savings will be counted or credited for all measures that have claimed savings. This Protocol contains requirements for the allowable methods for three types of evaluation studies: retention, degradation, and EUL analysis studies.

A persistence study measures changes in the net impacts that are achieved through installation/adoption of program-covered measures over time. These changes include retention and performance degradation. The definition of retention as used in this Protocol is the proportion of measures retained in place and that are operable. Effective useful life (EUL) is the estimate of the median number of years that the measures installed under the program are still in place and operable (retained).⁸⁰

The primary purpose of this Protocol is to provide ex-post estimates of effective useful life and performance degradation for those measures whose estimates are either highly uncertain and/or have not been covered in studies over the past 5 years. These results will be used to make prospective adjustments to the measure level EUL estimates and performance degradation estimates for Program Years 2009 and beyond, but will not be used for retroactive adjustments of the performance of the 2006-2008 portfolios.

The Effective Useful Life Evaluation Protocol is established to ensure that all persistence-related evaluations are conducted using evaluation methods deemed acceptable based upon the assigned level of rigor for that evaluation. The identification of allowable methods is one component of this Evaluation Protocol that helps ensure greater reliability in the energy and demand savings estimates from California's energy efficiency efforts. The Joint Staff can assign different levels of rigor for each measure in any study under this Protocol, thus allowing the Joint Staff the flexibility to allocate evaluation resources according to the needs of the Portfolio given uncertainties in the expected savings, the size of expected savings, the program budget and other criteria. The Joint Staff will instruct the evaluation contractors to use specific rigor levels based upon the Joint Staff's application of the decision criteria contained in this Protocol and evaluation resource allocations.

Rigor is defined as the level of expected reliability. The higher the level of rigor, the more confident we are that the results of the evaluation are both accurate *and* precise, i.e., reliable.

⁸⁰ These definitions are as in the Glossary. Some are the same and some have been modified from what was used in the prior M&E Protocols. There are, however, inconsistencies across different states in how these terms are used. Evaluators conducting EUL evaluations in California need to be familiar with the current California definitions.

That is, reliability and rigor are synonymous. Reliability is discussed in the Impact Evaluation Protocol, the Sampling and Uncertainty Protocol, and in the *Evaluation Framework* where it is noted that sampling precision does not equate to accuracy. Both precision and accuracy are important components in reliability, as used by the CPUC. Each evaluation study will be assigned a specific evaluation rigor level for its primary evaluation objectives to guarantee that a minimum standard is met.

Past experience presents a few important notes of caution. Many past persistence studies were unable to provide results that were significantly different (statistically) from the ex-ante results, so that most of the current ex-post EULs are the same as the ex-ante estimates. Besides finding relatively high retention rates in most cases, a consistent and important finding in these studies is that a longer period of time is needed for conducting these studies, so that larger samples of failures are available, and so that technology failure and removal rates can be better documented and used to make more accurate assessments of failure rate functions. The selection of what to measure, when the measurements should be launched, and how often they should be conducted are critical study planning considerations that Joint Staff will direct to ensure reliable results are achieved.

Performance degradation includes both (1) technical operational characteristics of the measures, including operating conditions and product design, and (2) human interaction components and behavioral measures. This Protocol refers to these two different components of performance degradation as technical degradation and behavioral degradation, respectively. (Performance degradation studies are also referred to in this Protocol more simply as degradation studies.)

Performance degradation accounts for both time-related and use-related change in the energy savings from an energy efficient measure or practice relative to a standard efficiency measure or practice. It is important to note that the energy savings over time is a difference rather than a straight measurement of the program equipment/behavior. It is the difference over time, between the energy usage of the efficient equipment/behavior and the standard equipment/behavior it replaced that is the focus of the measurement. Energy efficiency in both standard and high efficiency equipment often decreases over time. The energy savings over time is the difference between these two curves. The technical degradation factor is a set of ratios for each year after installation/adoption as the proportion of savings obtained in that year compared to the first-year savings estimate, regardless of the retention estimate or EUL (which is applied separately to obtain overall savings persisted). The technical (or behavioral) degradation factor could be 1.0 for each year in the forecast (often 20-year technical degradation factors are estimated) if the energy efficiency decreases (energy usage increases) by the same percentage each year as the standard equipment. This is the case where technical degradation rates are the same for both types of equipment. The technical (or behavioral) degradation factor would be higher if the efficient equipment holds its level of efficiency longer/better than the standard equipment⁸¹ and lower if there is more relative degradation.

Technical degradation studies may not be routinely required in the 2006-2008 round of EM&V studies because the incremental level of this type of degradation measured in five persistence

⁸¹ This was found to be the case in 3 of the 25 measures studied in the five persistence studies conducted under the prior M&E Protocols: residential d/x air-conditioning, residential refrigerators, and agricultural pumps.

studies from 1995 to 2000 was found to be insignificant for over 95% of measures.⁸² Joint Staff, however, may require a technical degradation study at their discretion. These may be needed based upon comments and findings within impact evaluations that discover potential issues with technical degradation, technologies not assessed in the five prior studies, changes in technology for the efficient or standard equipment, or for other reasons. For example, a technical degradation study may be desired for duct sealing which has not been previously studied.

The prior persistence studies included human interaction/behavior in the assessments made for the 25 measures examined. The importance of this may have been most prominent in the assessments of daylighting and energy management systems (EMS). The contribution of longer-term behavioral impacts on energy savings expectations over time could be an issue that still needs further examination, particularly given the greater emphasis being seen on more recent measures that include larger behavioral interaction components. The large human influence in the degradation factors found in EMS may also suggest periodic re-assessment or more field-based measurement studies. Measures that may need to be considered for behavioral degradation studies also include, but are not limited to, commissioning, retro-commissioning, operations and maintenance (O&M) efficiency efforts, programmable thermostats, and specific behavior-based initiatives. Joint Staff will select which measures will receive what types of studies and determine the scope of those evaluation studies focusing on behavioral degradation.

It is expected that evaluation contractors will respond to requests for proposals (RFPs) for retention, EUL, and performance degradation evaluations with proposals that meet the standards contained in this Protocol. The minimum allowable methods, sample criteria, and data collection criteria for these types of evaluations are provided later in this EUL Evaluation Protocol. In their proposal, evaluation contractors may propose (in addition to Protocol compliant methods) optional methods, if the contractor can clearly demonstrate that the optional methods provide at least as much rigor and accuracy as the Protocol-covered approach.

Audience and Responsible Actors

The audience and responsible actors for this Protocol include the following:

- Joint Staff evaluation planners – will use the Protocol to determine: (1) when special studies are needed for evaluating the retention, EUL, and degradation of particular measures, (2) as input into the RFPs for the evaluation of retention, EUL, and degradation, and (3) as background and criteria for use in reviewing retention, EUL, and degradation evaluation plans, managing the retention, EUL, and degradation evaluations, and reviewing retention, EUL, and degradation evaluation reports and results.
- Evaluation project team – will use the Protocol to make sure that the evaluations of the retention, EUL, and degradation of particular measures are based upon the level of rigor(s) designated by the Joint Staff. They will also use the Protocol to double-check that the

⁸² These five persistence studies conducted during the prior M&E Protocols and under supervision of the CADMAC Persistence Subcommittee are referred to as Persistence 1 (P1), Persistence 2 (P2), Persistence 3A (P3A), Persistence 3B (P3B), and Neg-TDF Supplement (PNg). These studies covered 25 measures and can be found in the CALMAC searchable database: 2023.pdf, 19980514CAD0006MR.pdf, 2028.pdf, 19990223CAD0003MR.pdf, and 19990223CAD0004MR.pdf.

Protocol requirements have been met as they conduct, complete, and report the retention, EUL, and degradation evaluations.

- Portfolio administrators – will use the Protocol to understand how the retention, EUL, and degradation evaluations will be conducted on their programs, and to understand the evaluation data needs to support the retention, EUL, and degradation evaluations.
- Program Implementers – will use the Protocol to understand how the retention, EUL, and degradation evaluations will be conducted on their programs. Often, they will be required to provide data to support the retention, EUL, and degradation evaluations.
- ISO / System planners – will utilize retention, EUL, and degradation estimates and uncertainty estimates for load forecasting and system planning.

Overview of the Protocol

This section briefly describes the three Protocols contained within the EUL Evaluation Protocol, Protocol rigor levels, key metrics assessed, and assessment inputs and outputs. This section is followed by sections that present the three Protocols that describe the allowable minimum methods for retention, degradation, and EUL studies. The reporting requirements for studies conducted within this Protocol are provided. This is followed by a short section providing guidance on the skills required by evaluators to conduct the type of studies described in this Protocol. The last section provides a brief summary list of the steps needed to comply with the EUL Evaluation Protocol.

Protocol Types

The Effective Useful Life Evaluation Protocol contains three Protocols, each providing the minimum requirements for: (1) retention studies, (2) degradation studies, and (3) EUL analysis studies. Each Protocol has two levels of rigor (Basic and Enhanced).

For each study, Joint Staff and their evaluators should examine opportunities for coordination (e.g., sampling and identifying and marking measures for further study) with impact studies: for example, it may be possible to conduct an analysis of retention for some measures in an impact study, to use the same sample, or use the same sample in a later study. But there are limitations in coordination: e.g., budget, appropriate sample, issues with ensuring random sampling across different objectives, etc. Strategically, it may be best to examine coordination for three different types of coordination: concurrent studies (e.g., examining current program impacts and retention from earlier participation at the same large commercial/industrial sites), past studies (i.e., using samples and information from prior impact studies for later retention or degradation studies), and future studies (e.g., collecting placement/location information or tagging in an impact study in order to assist a future retention study).

An example of how findings from the three types of persistence evaluations would work together to provide a measure's overall persistence is presented graphically in Figure 7.

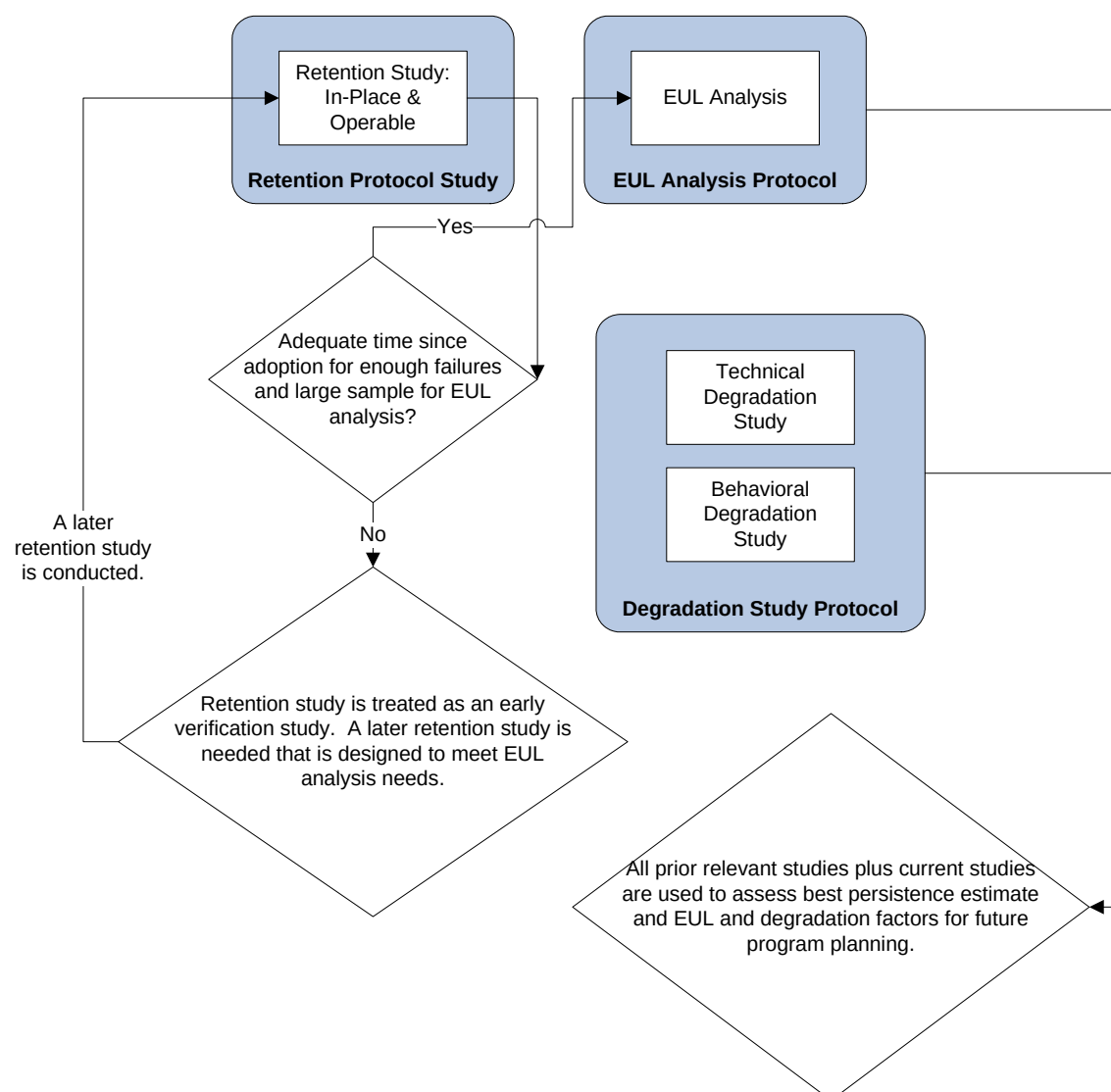


Figure 7. An Example of How Findings Across the Three Types of Studies Would Work Together for Persistence Evaluations

Rigor

When the Joint Staff decides a measure will receive a retention, EUL, and/or degradation evaluation, it also selects the level of evaluation rigor that is required. The Effective Useful Life Evaluation Protocol establishes the methods appropriate for the type of retention, EUL, and degradation evaluation designated to be conducted for the assigned level of evaluation rigor. In this way, the Protocol establishes a minimum level of evaluation rigor in order to ensure that the retention estimates, degradation factors, and EUL estimates produced are at the level of reliability needed to support the overall reliability of the savings in the administrator's Portfolio and the statewide Portfolio.

The level of rigor provides a class of allowable methods in order to offer flexibility for the potential evaluation contractors to assess, and propose the most accurate and cost-effective methods that meet the Joint Staff's needs. The principle is to provide minimum specifications for allowable methods, sample size criteria, and minimum data collection specifications and yet encourage evaluation contractors to utilize both the art and science of evaluation to develop affordable and quality evaluations that produce reliable savings estimates. There are two levels of rigor for each of the three Protocols: Basic and Enhanced, as shown in Figure 8. (The requirements for these rigor levels are described below).

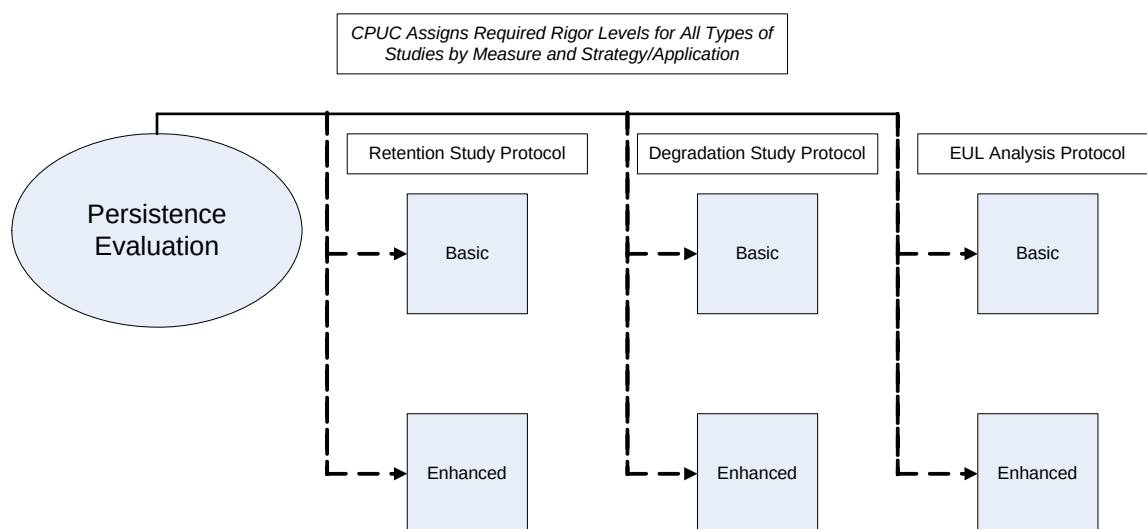


Figure 8. Protocols and Rigor Levels for EUL Evaluations

Joint Staff may assign rigor levels for evaluation studies covered in this Protocol for a measure or group of measures, for a measure within a delivery strategy, sector, or application. Separate retention estimates, degradation factors, and EULs may be required for measures by delivery strategy, sector, and application as assigned by Joint Staff. (Further discussion is provided below.)

Key Metrics, Inputs, and Outputs

Retention, EUL, and degradation evaluations will draw upon relevant data obtained from program databases, program descriptions, DEER database, work papers developed during program planning, on-site measurements, observational data, survey and interview data collection, manufacturers' studies, ASHRAE studies, laboratory studies, and other prior study data and reports. The use of these resources to support the planning and implementation of retention, EUL and degradation studies will help produce more reliable retention, degradation and EUL estimates.

Retention studies will provide the percent of the measures retained, along with clear descriptions of the methods used to determine measure-specific retention rates. In addition, these studies will provide complete definitions of what is considered an “operable condition” that constitutes a

retained status, and describe the testing criteria used to determine the operable status. Reporting of the retention estimates and degradation factors will also include a clear description of the methods employed and any adjustments made to ensure that the estimates appropriately represent all program installed/adopted measures without bias associated with changes in occupancy or location. The location where the measure was originally located needs to be maintained within a sample, regardless of occupant status, in cases where measures are not moved. For measures/behaviors that can be portable or easily moved, the study will verify the location and use of the measures and determine if they are still being used in a way that provides the projected savings. For the purposes of this Protocol, loss of retention is assumed when participants have moved out and taken the measures with them, unless the study provides reliable installation and energy savings use verification for the retaining of measures within the same utility service territory. (Finding and tracking movers for this purpose, however, is not required.)

Because EUL evaluations must follow the Sampling and Uncertainty Protocol, evaluators must also assess, minimize, and mitigate potential bias and present the achieved level of precision (including relative precision, error bounds, coefficients of variation, standard deviations, and error ratios) for interpreting information, summarizing retention estimates, degradation factors, and EULs and their precision by measure and strategy/application, and providing the information necessary for future evaluation planning. Where precision is calculated from chaining or pooling of evaluation study efforts, the above precision information should be provided for each study effort as well as the combined result.

All studies and evaluations conducted under this Protocol must comply with the reporting requirements contained in this Protocol.

Retention Study Protocol

These are minimum standard Protocols. All methods with higher rigor are allowable as they exceed the minimum criteria. For example, if the measure has a Joint Staff-assigned Basic Rigor and the method proposed by the evaluation contractor is an option under the Enhanced Rigor level, this method will be acceptable for meeting the Protocol if it meets budget and timing constraints. The Enhanced Rigor approach is the preferred approach for all retention studies. The Basic Rigor level may be assigned where this is more reasonable given the technology involved and budget constraints.

The Retention Study Protocol is summarized in Table 11. Further description, additional requirements, clarification, and examples of this Protocol are presented after the table. Being in compliance with the Retention Study Protocol means that the methods used and the way in which they are utilized and reported meet all the requirements discussed within this section (not just those within the summary table or those within the text) to provide unbiased reliable retention estimates. These Protocols sometimes reference other documents that provide examples of applicable methods. However, the operative requirements are only those stated in these Protocols, and not in the other references.

Measure retention studies collect data to determine the proportion of measures that are in place and operational. The primary evaluation components of a measure retention study are research design, survey-site visit instrument design, establishing the definition of an operable status

condition, identifying how this condition will be measured, and establishing the data collection and analysis approach. The measure retention estimate can be a straightforward calculation from the data collected. The key planning document associated with the study is the evaluation plan, which presents and discusses the methods for these components, as well as describing the data collection field efforts to be employed to support the data collection approach, and the study reporting to be delivered.

Joint Staff will decide which measures must receive retention studies and whether these studies must be conducted by delivery strategy, sector or other segmentation scheme in order to obtain reliable EUL estimates that can be used as a basis for future program planning. The evaluation contractor is expected to assess these instructions and work with Joint Staff to ensure that the most appropriate and cost-effective retention evaluation design is developed. This should be done as part of the initial evaluation planning and be completed prior to the completion of the final approved Evaluation Plan.

All retention evaluations are required to have a detailed evaluation plan. The evaluation plan needs to include a number of components to support an assessment of the adequacy and approach of the evaluation effort. These include the following components:

- Cover page containing the measures and delivery strategies or applications included in the retention evaluation, program names in the portfolios that include these, program administrators for these programs and their program tracking number(s), evaluation contractor, and the date of evaluation plan.
- Table of Contents.
- High-level summary overview of the measures and delivery strategies or applications included in the retention evaluation, the programs affected, and the evaluation efforts.
- Presentation of the evaluation goals and researchable issues addressed in the evaluation.
- Description of how the evaluation addresses the researchable issues, including a description of the evaluation priorities and the use of assigned rigor levels to address these priorities.
- A discussion of the reliability assessment to be conducted, including a discussion of the expected threats to validity, sources of bias, and a short description of the approaches planned to reduce threats, bias, and uncertainty.
- Task descriptions of the evaluation efforts.
- Review of any related retention and EUL study planning efforts prepared for Joint Staff to include prior estimation of failure sample size requirements, panel retention data needs and availability, and data tagging and collection efforts for these measures.
- Detailed description of the sampling rationale, methods, and sample sizes.
- Detailed description of the definition and methods for determining an operational condition.

- Detailed description of the information that will be needed from the Program Administrators in order to conduct the evaluation that will be included in evaluation-related data requests, including an estimate of date that the information will be needed or for which accessibility to the data is needed.
- Total evaluation budget and a task-level evaluation budget for the study; and
- Contact information for the evaluation manager, including, mail address, telephone numbers, fax numbers and e-mail address.

The evaluation plan should be written in a style and with enough detail that it can be clearly understood by program administrators, policy makers, and evaluation professionals.

Table 11. Required Protocols for Measure Retention Study

Rigor Level	Retention Evaluation Allowable Methods
Basic	1. In-place and operable status assessment based upon on-site inspections. Sampling must meet the Basic Rigor Level requirements discussed in this Protocol and must meet the requirements of the Sampling and Uncertainty Protocol. (The sampling requirements of this Protocol may need to meet the sampling requirements for the subsequent EUL study. See below specification.) 2. Non-site methods (such as telephone surveys/interviews, analysis of consumption data, or use of other data, e.g. from EMS systems) may be proposed but must be explicitly approved by Joint Staff through the evaluation planning process. Sampling must meet the Basic Rigor Level requirements discussed in this Protocol and must meet the requirements of the Sampling and Uncertainty Protocol. (The sampling requirements of this Protocol may need to meet the sampling requirements for the subsequent EUL study. See below specification.)
Enhanced	1. In-place and operable status assessment based upon on-site inspections. Sampling must meet the Enhanced Rigor Level requirements discussed in this Protocol and must meet the requirements of the Sampling and Uncertainty Protocol. (The sampling requirements of this Protocol may need to meet the sampling requirement for the subsequent EUL study. See below specification.)

The analysis of the retention data in either a Basic Rigor or Enhanced Rigor level study must include reporting the retention estimate as found in the study. Nevertheless, an assessment should be included as to whether one model or one brand showed a strong affect on the retention estimate where the exclusion of this model or brand from programs would change the expected resulting EUL by more than 25 percent (25%). If this is suspected prior to completion of the Evaluation Plan, the retention study sampling design and sample sizes may need to be conducted to produce retention estimates for this model or brand separately from that of other models/brands where both retention estimates meet the sampling and precision criteria required for the assigned level of rigor for that retention study. Joint Staff and their evaluators may want to consider whether the retention study sample should be part of an impact evaluation sample, or have separate samples for retention and impacts studies.

Basic Rigor

In-place and operable status assessment based upon on-site inspections is considered the default requirement for retention studies.

The in-place assessment shall be verified through on-site inspections of facilities. Measure, make and model number data shall be collected and compared to participant program records as applicable. As-built construction documents may be used to verify selected measures where access is difficult or impossible (such as wall insulation). Spot measurements may be used to supplement visual inspections, such as solar transmission measurements and low-e coating detection instruments, to verify the optical properties of windows and glazing systems.

Correct measure operation shall be observed and compared to project design intent. Often this observation is a simple test of whether the equipment is running or can be turned on. This can also include, however, changes in application or sector such that the operational nature of the equipment no longer meets project design intent. For example, working gas-cooking equipment that had been installed in a restaurant but is now installed in the restaurant owner's home is most likely no longer generating the expected energy savings and would not be counted as an program-induced operable condition.

Non-site methods (telephone surveys/interviews, analysis of consumption data, or use of other data, e.g. from EMS systems) may be proposed along with a detailed description as to why the proposed method(s) would be reliable for the study measures in their strategies/applications based upon theoretical and past study justifications. All methods, however, must be assessed and approved by Joint Staff through the evaluation planning process and explicit acceptance of proposed non-site methods are required prior to their being used.

The reasons for lack of retention, and the rates of non-retention, should be gathered when feasible for use in developing EUL study designs and future retention studies.

In most cases, there will be a sample size requirement for an EUL study that will be used to determine the sample size requirement for a corresponding retention study since a survival analysis will be based on data collected earlier in a corresponding retention study. Thus, for a given retention time period under study, the sample size for a retention study must meet any prior sample size requirements determined for EUL studies on the proposed measures in these strategies/applications and must meet the requirements of the Sampling and Uncertainty Protocol. However, there are two conditions that could arise that should be addressed:

- If there is no *EUL*-determined required sample size for the retention time period under study and the study retention period is within 30% of the expected EUL, then the sample size required for an EUL study must be calculated in order to determine the retention study sample size requirement. This includes using power analysis at a power of at least 0.7 for the Basic Rigor level to determine the sample size required at a 90% confidence level (alpha set at 0.10), and then deriving the required retention sample size based upon the proportion of the original pool expected to be found in-place and operable (the *ex-ante* EUL). (See the EUL Analysis Protocol and Appendix D for more information concerning the use of power analysis for determining sample size requirements.)

- If there is no *EUL*-determined required sample size for the retention time period under study and the study retention period is not within 30% of the expected EUL, then the retention study sample size requirement should be based upon the coefficient of variation, standard deviation, and other available estimates of variance for the percent of equipment that is in place and operable from prior studies. The sample size should be large enough to attain a minimum of 30 % precision at the 90% confidence level.

Enhanced Rigor

The in-place assessment shall be verified through on-site inspections of facilities. Measure, make and model number data shall be collected and compared to participant program records as applicable. As-built construction documents may be used to verify measures such as wall insulation where access is difficult or impossible. Spot measurements may be used to supplement visual inspections, such as solar transmission measurements and low-e coating detection instruments, to verify the optical properties of windows and glazing systems.

Correct measure operation shall be observed and compared to project design intent. Commissioning reports (as applicable) shall be obtained and reviewed to verify proper operation of installed systems. If measures have not been commissioned, measure design intent shall be established from program records and/or construction documents; and functional performance testing shall be conducted to verify operation of systems in accordance with design intent. This must also include an assessment of whether changes in application or sector are such that the operational nature of the equipment no longer meets project design intent. For example, working gas-cooking equipment that had been in a restaurant but is now in the restaurant owner's home is no longer meeting project design intent and is no longer generating the expected energy savings.

Analysis of consumption data or use of data from EMS systems may be proposed along with a detailed description as to why the proposed method(s) would be reliable for the study measures in their strategies/applications based upon theoretical and past study justifications. (Telephone surveying or interviewing techniques are not presented as an allowed approach within the Enhanced Rigor level for retention studies.) All methods, however, must be assessed and approved by Joint Staff through the evaluation planning process and explicit acceptance of proposed non-site methods are required prior to their being used.

The reasons for lack of retention, and the rates of these, should be gathered when feasible for use in developing EUL study designs and future retention studies. For example, in one study, the removal rate of refrigerators during the first five years was found to be higher for locations where the consumer moved.⁸³ It could be expected that as the refrigerator ages the probability for older refrigerators being moved with consumers may decrease. As a result, an improved EUL function would reflect the risk of participants moving with their refrigerators.

⁸³ This hypothesis, its testing, and consequences were examined in a study by Quantum Consulting and Megdal & Associates in the *Retention Study of Pacific Gas and Electric Company's 1996 and 1997 Appliance Energy Efficiency Programs*, Study ID 373 1R1, March 2001.

In most cases, there will be a sample size requirement for an EUL study that will be used to determine the sample size requirement for a corresponding retention study since a survival analysis will be based on data collected earlier in a corresponding retention study. Thus, for a given retention time period under study, the sample size for a retention study must meet any prior sample size requirements determined for EUL studies on the proposed measures in these strategies/applications and must meet the requirements of the Sampling and Uncertainty Protocol. However, there are two conditions that could arise that should be addressed::

- If there is no *EUL*-determined required sample size for the retention time period under study and the study retention period is within 30% of the expected EUL, then the sample size required for an EUL study must be calculated in order to determine the retention study sample size requirement. This includes using power analysis at a power of at least 0.8 for the Enhanced Rigor level to determine the failure sample size required at a 90% confidence level (alpha set at 0.10) and then deriving the required retention sample size based upon the proportion of the original pool expected to be found in-place and operational. (See the EUL Analysis Protocol and Appendix D for more information concerning the use of power analysis for determining sample size requirements.)

If there is no *EUL*-determined required sample size for the retention time period under study and the study retention period is not within 30% of the expected EUL, then the retention study sample size requirement should be determined based upon the coefficient of variation, standard deviation and other available estimates of variance for the percent of equipment that is in place and operable from prior studies. The sample size should be large enough to attain a minimum of 10 % precision at the 90% confidence level.

Degradation Study Protocol

These are minimum standard Protocols. All methods with higher rigor are allowable as they exceed the minimum criteria. For example, if the measure has been assigned a Basic level of rigor by the Joint Staff and the method proposed by the evaluation contractor is an option under Enhanced, this method will be acceptable for meeting the Protocol. The Enhanced Rigor approach is the preferred approach for all retention studies. The Basic Rigor level may be assigned as is reasonable given the technology involved and budget constraints.

The Degradation Study Protocol is summarized in Table 12. Further description, additional requirements, clarification, and examples of this Protocol are presented after the table. Being in compliance with the Degradation Study Protocol means that the methods used and the way in which they are utilized and reported meet all the requirements discussed within this section (not just those within the summary table or those within the text) to provide unbiased reliable estimates of the technical degradation factor. The Protocols sometimes reference other documents that provide examples of applicable methods. The requirements, however, are always those stated in these Protocols, which take precedence over all others in all circumstances.

Performance degradation studies produce a factor that is a multiplier used to account for both time-related and use-related change in the energy savings of a high efficiency measure or practice relative to a standard efficiency measure or practice. It is important to note that the degradation study is a relative difference measurement between the high efficiency

equipment/behavior and the non-high efficiency equipment/behavior over time (and not the relative level of retention). Said in a different way, it is the difference over time between the energy usage of the energy efficient equipment/behavior and the standard equipment/behavior it replaced. Appropriate standard measure comparisons are critical.

Studies must designate and clearly describe all the elements that will be analyzed for the degradation factor produced. If only a technical degradation factor is produced, or only a behavioral degradation factor is produced for a measure that contains both (i.e. when a technical degradation factor and a behavioral degradation factor is associated with the same piece of equipment), this must be clearly noted. If the equipment has both a technical and behavioral degradation factor, and one of these is to come from a previous study or another source, and the other is being addressed in the Protocol-covered study, both factors must be presented in the report, and the analysis must produce a combined factor for that equipment. If only one factor is being produced, the study must clearly describe whether this covers the full degradation factor to be used for that measure or indicate if additional research is needed to establish a more complete/reliable factor that includes both technical and behavioral degradation. For example, a measure that is purely behavioral (e.g., maintenance behavior schedules) would only receive a behavioral degradation factor and this should be clearly described, however, a measure such as a programmable thermostat would need a combined performance degradation factor that incorporates the technical degradation factor for the measure and the behavioral degradation factor associated with the use of the measure as an energy saving device.

Since the degradation factor is the difference between the standard equipment/behavior and that of the program measures, the over time, changes in usage for standard and efficiency must be clearly assessed/measured and reported or the component differences and their changes in usage over time must be explained and assessed/measured.

All degradation evaluations are required to have a detailed evaluation plan. The evaluation plan needs to include a number of components to support an assessment of the adequacy and approach of the evaluation effort. These include the following components:

- Cover page containing the measures and delivery strategies or applications included in the degradation evaluation, program names in the portfolios that include these, program administrators for these programs and their program tracking number(s), evaluation contractor, and the date of evaluation plan.
- Table of Contents.
- High-level summary overview of the measures and delivery strategies or applications included in the degradation evaluation, the programs affected, and the evaluation efforts.
- Presentation of the evaluation goals and researchable issues addressed in the evaluation.
- Description of how the evaluation addresses the researchable issues, including a description of the evaluation priorities and the use of assigned rigor levels to address these priorities.

- A discussion of the reliability assessment to be conducted, including a discussion of the expected threats to validity, sources of bias, and a short description of the approaches planned to reduce threats, bias, and uncertainty.
- Task descriptions of the evaluation efforts.
- Detailed description of the sampling rationale, methods, and sample sizes.
- Detailed description of the methodology to be used for the assessment for both the standard equipment/behavior and that for the efficient equipment/behavior, and the approach to be used for quantifying the difference between these two conditions. This condition applies regardless if the study is determining the EUL by assessing the equipment as a whole unit, or if the assessment is conducted for a key component of the equipment, or if the assessment is based on engineering assumptions about the expected performance or performance life of a component of the equipment.
- Detailed description of the information that will be needed from the Program Administrators in order to conduct the evaluation that will be included in evaluation-related data requests, including an estimate of date that the information will be needed or for which accessibility to the data is needed.
- Total evaluation budget and a task-level evaluation budget for the study; and
- Contact information for the evaluation manager, including, mail address, telephone numbers, fax numbers and e-mail address.

The evaluation plan should be written in a style and with enough detail that it can be clearly understood by program administrators, policy makers, and evaluation professionals.

Table 12. Required Protocols for Degradation Study

Rigor Level	Allowable Methods for Degradation Studies
Basic	1. Literature review required for technical degradation studies across a range of engineering-based literature, to include but not limited to manufacturer's studies, ASHRAE studies, and laboratory studies. Review of technology assessments. Assessments using simple engineering models for technology components and which examine key input variables and uncertainty factors affecting technical degradation. 2. Telephone surveys/interviews with a research design that meets accepted social science behavioral research expectations for behavioral degradation.
Enhanced	1. For technical degradation: field measurement testing. 2. For behavioral degradation: field observations and measurement.

Basic Rigor

Technical degradation studies require a literature review for the measures under study. The literature search should include journal articles, conference proceedings, manufacturer publications, publications of engineering societies (e.g., ASHRAE), national laboratories, and government agencies, and the gray literature (i.e., studies that are not widely published but are available upon request). In addition, technology assessments using simple engineering models for technology components will be conducted. Studies will be conducted that examine key input variables and uncertainty factors affecting technical degradation.

Laboratory testing may be used to determine the technical degradation factor(s) for the Basic Rigor level. Laboratory testing involves the measurement of energy use of both energy efficient and standard equipment over time, but in unoccupied facilities. Laboratory testing must account for the operational conditions expected for installations obtained through the California programs that incorporate the measure in their service mix.

Telephone surveys/interviews with a research design that meets accepted social science behavioral research expectations for behavioral degradation. The use of the term acceptable, in this case, means that the approach and the data collection methods would pass a peer review process using highly experienced professional social science researchers, in such a way that peers would support and defend the approach as being objective and reliable within the ability of the approach selected. The types of questions asked will focus on whether the energy efficient and standard equipment are being used/operated as designed and the reasons for their non-use or changes in use.

Enhanced Rigor

Technical degradation studies at the Enhanced Rigor level require field measurement. Field measurement involves the measurement of energy use for both the energy efficient and the standard equipment over time – these measures would be located in occupied facilities. These measurements must be designed to collect data on the equipment or equipment components in order to reduce the greatest uncertainties associated with the degradation factor estimates and be conducted with sample sizes to allow for 90% confidence and 30% precision in these measurements.

Behavioral degradation studies focus on the observation (and measurement, if applicable) of the use of energy efficient and standard equipment in facilities. The studies can be short term (one-time site visits) or long term (periodic site visits) to assess if the measures are being used as designed. The types of questions asked will focus on whether the energy efficient and standard equipment are being used/operated as designed and the reasons for their non-use or changes in use. The self-reports will be matched with observational data for confirmation. Measurement studies must be designed to collect data to reduce the greatest uncertainties associated with the degradation factor estimates and be conducted on sample sizes to allow for 90% confidence and 30% precision in these measurements.

Effective Useful Life Analysis Protocol

These are minimum standard Effective Useful Life (EUL) Protocols. All methods with higher rigor are allowable as they exceed the minimum criteria. For example, if the measure has a Joint Staff assigned rigor of Basic and the method proposed by the evaluation contractor is an option under Enhanced, this method will be acceptable for meeting the Protocol.

The EUL Analysis Protocol is summarized in Table 13. Further description, additional requirements, clarification, and examples of this Protocol are presented after the table. Being in compliance with the EUL Analysis Protocol means that the methods used and the way in which they are utilized and reported meet all the requirements discussed within this section (not just those within the summary table or those within the text) to provide unbiased reliable estimates of EUL. The Protocols sometimes reference other documents that provide examples of applicable methods. The requirements, however, are always those stated in these Protocols, which take precedence over all others in all circumstances.

The objective of the EUL analysis studies is to estimate the ex-post EUL, defined as the estimate of the median number of years that the measures installed under the program are still in place, operable, and providing savings. Evaluators are expected to develop a plan for estimating survival functions for measures included in the scope of their work. The plan should incorporate an assessment of the study's ex-post EUL compared to the findings of other studies on this measure for this delivery strategy/application and to the EUL contained in the DEER database (if one exists). The study should also provide recommendations for whether the new EUL should be substituted for the existing ex-ante EUL for future program planning or if another DEER category for the measure should be developed. The EUL studies are also required to report the findings from the most recent degradation studies related to the EUL study measures/applications, so that the EUL report is a depository for all current persistence studies for the study measures/applications. This will assist Joint Staff and future evaluators find all relevant persistence information for a measure/application in one location.

Joint Staff will decide which measures must receive EUL studies and whether these studies must be conducted by delivery strategy, sector, or other segmentation scheme in order to obtain reliable EUL estimates that can be appropriately used for various future program planning alternatives. The evaluation contractor is expected to assess these instructions and work with Joint Staff to ensure that the most appropriate and cost-effective evaluation design is developed, so reliable EUL estimates are obtained to meet this purpose. This should be done as part of initial evaluation planning and be completed prior to completion of the Evaluation Plan.

As part of the evaluation plan, the evaluator should propose a method for estimating a survival function and a survival rate of measures installed over time using standard techniques (see below). This should include an identification of factors that might lead to lower survival rates and a discussion of how the confidence intervals for the survival functions derived will be estimated. This should also include a discussion of potential sources of bias in the methods proposed and how these sources of bias will be mitigated.

All EUL analysis evaluations are required to have a detailed evaluation plan. The evaluation plan needs to include a number of components to support an assessment of the adequacy and approach of the evaluation effort. These include the following components:

- Cover page containing the measures and delivery strategies or applications included in the EUL analysis evaluation, program names in the current and past (up to 10 years previous programs) portfolios that include these, program administrators for these

programs and their program tracking number(s) if applicable, evaluation contractor for those programs, and the date of evaluation plan.

- Table of Contents.
- High-level summary overview of the measures and delivery strategies or applications included in the evaluation, the programs affected, and the evaluation efforts conducted.
- Presentation of the evaluation goals and researchable issues addressed in the current EUL evaluation being planned.
- Description of how the evaluation addresses the researchable issues, including a description of the evaluation priorities and the use of assigned rigor levels to address these priorities.
- A discussion of the reliability assessment to be conducted, including a discussion of the expected threats to validity, sources of bias, and a short description of the approaches planned to reduce threats, bias, and uncertainty.
- Task descriptions of the evaluation efforts.
- Review of any related EUL study planning efforts prepared for Joint Staff to include prior estimation of failure sample size requirements, panel retention data needs and availability, and data tagging and collection efforts for these measures.
- Detailed examination of related retention studies, assessment of prior and concurrent retention studies and recommendation of what additional data must be collected and, if so, why. If additional data collection is proposed, then a detailed description of the sampling rationale, methods, and sample sizes must be included.
- Detailed description of the information that will be needed from the Program Administrators in order to conduct the evaluation that will be included in evaluation-related data requests, including an estimate of date that the information will be needed or for which accessibility to the data is needed.
- Total evaluation budget and a task-level evaluation budget for the study; and
- Contact information for the evaluation manager, including, mail address, telephone numbers, fax numbers and e-mail address.

The evaluation plan should be written in a style and with enough detail that it can be clearly understood by program administrators, policy makers, and evaluation professionals.

As noted below, we rely on power analysis for helping to differentiate Basic Rigor from Enhanced Rigor (see Appendix D for more discussion and references). Statistical power is the probability that statistical significance will be attained, given that there is a measurable treatment effect. Power analysis is a statistical technique that can be used (among other things) to determine sample size requirements to ensure that statistical significance can be found. Power analysis is a required component in the Protocol to assist in determining required sample sizes.

The Basic level of rigor in the EUL Protocols requires that a 0.70 level of power be planned at the 90% level of confidence. While the Enhanced level of rigor requires that a 0.80 level of power be planned also at the 90% level of confidence. In determining sample sizes in the

research planning process, values for key parameters can be varied in an attempt to balance a level of statistical power, the alpha, the duration of the study, and the effect size, all determined with an eye on the budget constraints. The values will probably be unique to each measure selected for study. The differing power requirements between the Basic and Enhanced level of rigor drives up the required sample size to meet the Enhanced Rigor level versus that needed to meet the Basic Rigor level. The results of the power analysis will be combined with professional judgment and past studies to arrive at the required sample sizes. The selected sample size, the results of the power analysis, and the justification for the sample size proposed must be included in the evaluation plan. This evaluation plan must be approved by Joint Staff prior to sample data collection.

Table 13. Required Protocols for EUL Analysis Studies

Rigor Level	Allowable Methods for EUL Analysis Studies
Basic	Classic survival analysis (defined below) or other analysis methods that specifically control for right-censored data (those cases of failure that might take place some time after data are collected) must be attempted. For methods not accounting for right-censored data, the functional form of the model used to estimate EUL ("model functional form") must be justified and theoretically supported. Sampling must meet the Basic Rigor Level requirements discussed in this Protocol and must meet the requirements of the Sampling and Uncertainty Protocol. Sample size requirements will be determined through the use of power analysis, results from prior studies on similar programs, and professional judgment. Power analysis used to determine the required sample size must be calculated by setting <u>power to at least at 0.7</u> to determine the sample size required at a 90% confidence level (alpha set at 0.10). Where other analyses or combined functional forms are used, power analysis should be set at these parameters to determine required sample sizes for regression-based approaches and a 90% confidence level with <u>30% precision</u> is to be used for non-regression components.
Enhanced	Classic survival analysis (defined below) or other analysis methods that specifically control for right-censored data (those cases of failure that might take place some time after data are collected) must be attempted. The functional form of the model used to estimate EUL ("model functional form") must be justified and theoretically supported. Sampling must meet the Enhanced Rigor Level requirements discussed in this Protocol and must meet the requirements of the Sampling and Uncertainty Protocol. Sample size requirements will be determined through the use of power analysis, results from prior studies on similar programs, and professional judgment. Power analysis used will set <u>power to at least to 0.8</u> to determine the sample size required at a 90% confidence level (alpha set at 0.10). Where other analyses or combined functional forms are used, power analysis should be set at these parameters to determine required sample sizes for regression-based approaches and a 90% confidence level with <u>10% precision</u> is to be used for non-regression components.

Basic Rigor

Current ex-ante EULs were developed using engineering experience and assumptions, past M&V-related evaluation efforts, and past EUL studies. Engineering analysis and M&V

observations suggest that energy efficiency measures generally last a certain average length of time and then rapidly move out of use as the measures reach their end of life service. However, these approaches have generally not considered retention and behavioral degradation in establishing the EUL estimates. Similarly, a few measures may continue to last significantly beyond their expected lifetime.

An initial approximation for most types of EUL forecasts efforts involve some form of a linear estimate, even if the estimate is not linear during the first years of use, or during the later years. This typically involves trying to fit a line to the observed data and use this to predict EUL estimates. Yet, the engineering experience for efficiency measures suggests that a linear model may not represent the survival function of many energy efficiency measures.

Common alternative models include logistic and exponential models. A variation of the logistic function can be used to describe a pattern of little loss in the early years with increasing loss as the measure approaches its expected life, with a flattening loss occurring thereafter.

The standard cumulative logistic probability function is:

$$P_i = F(Z_i) = F(\alpha + \beta X_i) = 1/(1 + e^{-(\alpha + \beta X_i)})$$

The logistic model is generally used to measure and predict probabilities that an event will occur. This model limits the end points to zero and one. The cumulative logistic, the logistic model, looks like the curve shown in Figure 9.

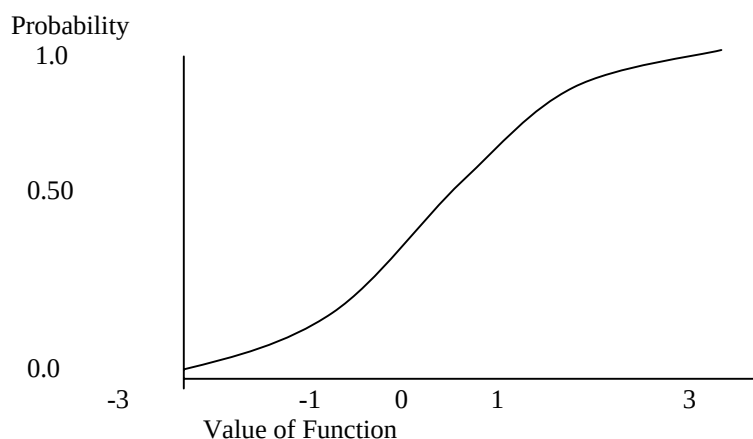


Figure 9. Cumulative Logistic Function

The logistic function that best fits the engineering observations described above relies upon a logistic function of time for identifying the EUL. This is:

$$F(Z_i) = 1 - [1/(1 + e^{-(t + \text{EUL})b})]$$

With the survival function as in Figure 10.

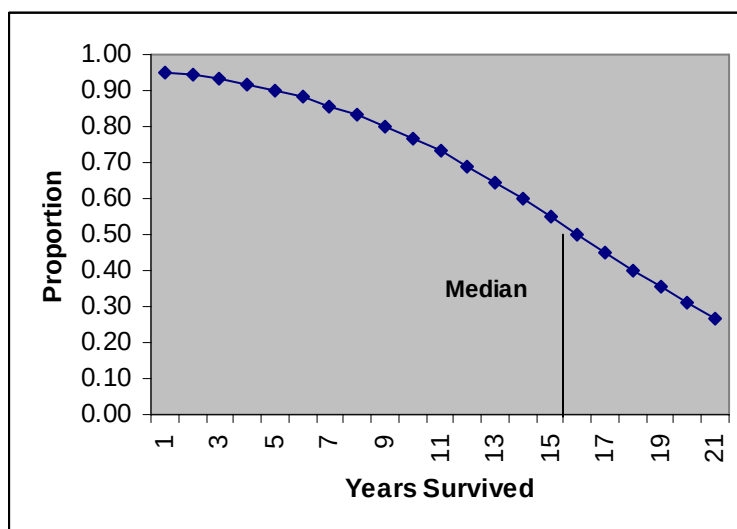


Figure 10. Logistic Survival Function with EUL=15 and $b=0.2$

Many energy efficiency retention studies examine energy efficiency equipment as being either there or not. This dichotomous scale allows the possibility of using classical survival analysis techniques. These outcomes are dichotomous, that is, they either occur or not and can be measured as zero or one events.

Classic survival analysis is specifically designed to account for the fact that “failures” might take place some time after when data are measured. In the statistical literature, these cases are said to be “right censored” – their failures are not included in the analysis, because the time period was not long enough to include their eventual failure. As a result, estimating the mean or median when data are right censored can provide a biased estimate. Classic survival analysis techniques have been developed that account for this right censorship in the data and are able to provide unbiased estimates. Thus, given enough data, many functional forms of survival analysis models (“model functional forms”) can be tested with available survival analysis statistical programs. The regression techniques available allow consideration of right censored data and can handle continuous time data, discrete time data, and other types of data.⁸⁴

⁸⁴ Multiple statistical modeling packages (SAS®, Stata®, SPSS®, R®, S+®, and others) provide survival analysis programs. There are several commercial and graduate textbooks in biostatistics that are excellent references for classic survival analysis. One of these used as reference for some of the prior EUL studies in California is the SAS® statistical package and the reference *Survival Analysis Using the SAS® System: A Practical Guide* by Dr. Paul D. Allison, SAS® Institute, 1995. Several model functional forms are available and should be considered for testing. These forms include logistic, logistic with duration squared (to fit expected pattern of inflection point slowing of retention losses), log normal, exponential, Weibull, and gamma. A few of many possible references include: *The Statistical Analysis of Failure Time Data* (Wiley Series in Probability and Statistics) by John D. Kalbfleisch, Ross L. Prentice, Wiley, 2003; *Survival Analysis: A Self-Learning Text* by David G. Kleinbaum, Mitchel Klein, Springer-Verlag New York, LLC, 2005; *Survival Analysis* by David Machin, Wiley, 2006; and *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence* by Judith D. Singer, John B. Willett, Oxford University Press, 2003.

Given the advantages of classic survival analysis for producing unbiased estimates of the EUL, both levels of rigor require attempting this method where applicable. The high demand for failure data and the need for differentiated data over time to best approximate a reasonable functional form have often meant that these models either did not provide algorithm convergence or produce reasonable results. Accordingly, other methods that specifically control for right censored data not yet well defined or explained must be attempted in its place, if they are appropriate to the hypothesized hazard functional form.

Where a specific method that controls for the issues associated with right censored data cannot be made workable or to produce reasonable results that can be justified, then the evaluator may resort to other methods to estimate the EUL. Nevertheless, the EUL estimate from other methods must either adjust or, at a minimum, discuss the likely potential bias in the EUL estimate given the inability to control for right censored data issues.

Sample size requirements will be determined through the use of power analysis, results from prior studies on similar programs, and professional judgment. Power analysis will set power to at least to 0.7 to determine the sample size required at a 90% confidence level (alpha set at 0.10) and then derive the required retention sample size based upon the proportion of the original pool expected to be found in-place and operable (ex-ante EUL) and the desired effect size as determine by Joint Staff. Where other analyses or combined functional forms are used, power analysis should be set at these parameters to determine required sample sizes for regression-based approaches and a 90% confidence level with 30% precision is to be used for non-regression components. Sampling and reporting of sampling and uncertainty must meet the requirements of the Sampling and Uncertainty Protocol.

Joint Staff may want to consider whether the EUL study sample should be part of an impact evaluation sample, or have separate samples for EUL and impacts studies. Joint Staff may assign separate retention and EUL analysis studies or joint studies as they find appropriate for the timing of the evaluations and efficiencies between studies.

Enhanced Rigor

All of the analysis requirements for the Basic Rigor level apply to the Enhanced Rigor level.

Sample size requirements will be determined through the use of power analysis, results from prior studies on similar programs, and professional judgment. Power analysis will set power to at least to 0.8 to determine the sample size required at a 90% confidence level (alpha set at 0.10) and then deriving the required retention sample size based upon the proportion of the original pool expected to be found in-place and operable (ex-ante EUL) and the desired effect size as determine by Joint Staff. Where other analyses or combined functional forms are used, power analysis should be set at these parameters to determine required sample sizes for regression-based approaches and a 90% confidence level with 10% precision is to be used for non-regression components. Sampling and reporting of sampling and uncertainty must meet the requirements of the Sampling and Uncertainty Protocol.

Joint Staff may want to consider whether the EUL study sample should be part of an impact evaluation sample, or have separate samples for EUL and impacts studies. Joint Staff may

assign separate retention and EUL analysis studies or joint studies as they find appropriate for the timing of the evaluations and efficiencies between studies.

Reporting Requirements

All EUL evaluations are expected to assess and discuss the differences between the (a) ex-ante EUL estimates from DEER or as otherwise approved by the Joint Staff and (b) the ex-post EUL estimates produced by the EUL evaluation study(ies). To the extent that the data gathered and evaluation analyses conducted can explain the causes for these differences, this must be presented and discussed. The evaluation report should note situations in which explanations are not possible due to lack of sufficient data or problems with interpretation. The EUL evaluation report must also include a recommendation of the EUL for the measure and delivery strategy/application that should be used for future program planning. This recommendation may take the form of recommending the replacement of a DEER EUL or the establishment of a new DEER category.

The EUL studies are also required to report the findings from the most recent degradation studies related to the EUL study measures/applications, so that the EUL report is a depository for all current persistence studies for the study measures/applications. This will assist Joint Staff and future evaluators find all relevant persistence information for a measure/application in one location.

All **reporting under this Protocol** should include the following:

1. Cover page containing the measures and delivery strategies or applications included in the retention evaluation, program names in the portfolios over the last 5 years that include these, program administrators for these programs and their program tracking number(s), evaluation contractor, and the date of evaluation plan.
2. Table of Contents.
3. High-level summary overview of the measures and delivery strategies or applications included in the evaluation, the programs affected, and the evaluation efforts.
4. Presentation of the evaluation goals and researchable issues addressed in the evaluation.
5. Description of how the evaluation addresses the researchable issues, including a description of the evaluation priorities and the use of assigned rigor levels to address these priorities.
6. Detailed description of the data collection and analysis methodology.
7. Current and prior retention results for selected measures given delivery strategy/application and their precision levels at a 90% confidence interval.
8. Retention, degradation, and EUL findings as is appropriate for the study assigned.
9. A discussion of the reliability assessment to be conducted, including a discussion of the expected threats to validity, sources of bias, and a short description of the approaches planned to reduce threats, bias, and uncertainty.
10. Contact information for the evaluation manager, including, mail address, telephone numbers, fax numbers and e-mail address.

In addition to the above, **retention studies** must also include the following:

- Description of initial and final sample of measures still surviving.
- Describe any findings on factors leading to the higher or lower retention rates.
- Description of removal reasons, their distribution, and potential issues created by different removal reasons and the research design and functional forms that should be investigated in future EUL studies for these measures.

In addition to the overall EUL study reporting requirements, **degradation studies** must also include the following:

- Describe any findings on factors leading to the relative degradation rates and absolute degradation rates, if available.
- Describe the impact of degradation on energy savings

In addition to the overall EUL study reporting requirements, **EUL analysis studies** must also include the following:

- Specific equations for survival functions and estimated precision of curve fit.
- Analysis of the ex-post EUL compared to the ex-ante EUL and comparison of to the methods and results from any prior retention, degradation, or EUL studies available for that measure (to include comparisons by delivery strategy and application).
- Recommended EUL for the measure and delivery strategy/application that should be used for future program planning.

Study Selection and Timing

A significant amount of funding has been spent in California conducting EUL studies under the prior M&E Protocols. These studies, completed through 2004, have been reviewed in a recent study (Skumatz et al. 2005).⁸⁵ Those measures with useful lives that have been confirmed in the last five years are less likely in need of additional study in the 2006- 2008 study period. Early work by Joint Staff in the EUL planning effort may include an initial study of required EUL sample sizes, review of prior EUL studies and their data collection methods, and an assessment of which measures should be prioritized for which types of studies. Important questions for early EUL planning efforts include the following:

- Which measures should obtain early panel data collection plans?
- Where can data be collected in the future through periodic retention studies that incorporate estimated removal dates prior to the retention study date (prospective studies that use retrospective methods in their site or telephone surveying)? For example, it may

⁸⁵ *Revised/Updated EULs based on retention and persistence studies results*, Revised Report, Lisa Skumatz, John Gardner, Skumatz Economic Research Associates, Inc., Superior, CO, 2005. It should be noted that several ninth-year studies have been or are being conducted since the summary report was produced.

be possible to conduct studies on data collected from past years on measures promoted in previous programs for conducting a retention study, or one can simply conduct retention studies on measures installed in the current programs.

- Which long-life measures should have a retrospective analysis conducted for obtaining an EUL estimate?⁸⁶ For example, a measure that was included in a previous program currently has an ex-ante EUL of over 7 years, so that obtaining a reliable ex-post EUL estimate in the next three years from a prospective approach would be highly unlikely. However, an ex-post EUL estimate may be obtainable by conducting retrospective analyses on the prior program sites.

The analysis could also include an assessment of the costs, benefits, and removal risks for tagging⁸⁷ and/or using radio frequency identification (RFID) chips by program implementers to simplify future retention studies. This information can be used to estimate study costs, timing of study RFPs, and as input into any risk analysis used to allocate resources for EUL studies.

The general rules for how often and what EUL/retention/degradation evaluations need to be conducted are determined by the Joint Staff. The Joint Staff will decide for each measure, if and when the measure will receive a retention study, a degradation study, and/or an EUL evaluation. They will also decide whether the studies need to be conducted for the measure in a single classification or segregated by delivery strategy or application and whether degradation studies will be overall or technology-based or behavior-based. A few examples of when this might occur are: (1) where early removal is a risk due to performance, comfort or aesthetic concerns, (2) when more detailed evaluation information is needed for a measure for future program planning, (3) to support an update of DEER, or (4) a new measure is being piloted or expanded in its use. Joint Staff may assess these situations through risk analysis to determine which types of EUL studies to undertake and when.

Guidance on Skills Required to Conduct Retention, EUL, and Technical Degradation Evaluations

The senior, advisory and leadership personnel for EUL analysis evaluation efforts need to have the specific skills and experience in regression and statistics proving an ability to be able to conduct classic survival analysis and handle EUL functional form and issue analysis, as well as the time budgeted for responsible project task leadership and quality control.

⁸⁶ The retention/EUL studies conducted under the prior M&E Protocols included panel studies, one-retention point site visit studies, site and telephone surveys that had field observation estimate, and consumer estimates on prior removal dates. The one-point retention studies generally provided only two time points for the retention analysis. This does not allow for information to help determine the appropriate functional form or for inflection points in removal rates. Panel studies offer the most reliable information but are quite expensive. Periodic site and telephone surveys with consumer estimates of removal dates offer the most cost-effective data if their reliability is sufficient for accurate EUL estimation. Some measures can have retention (in place and operational) reliably measured through telephone surveys (e.g., attic and wall insulation seeking remodeling occurrences) while others may require site visit verification and measurement.

⁸⁷ A 1994 study by ASW Engineering and KVD Research Consulting concluded that tagging equipment might not be viable due to retention issues with the tags.

There are methods that could employ significant primary survey or interview data collection. The evaluators using these methods should have sufficient experience implementing surveys, interviews, group interviews, and other types of primary data collection activities as are being recommended.

Engineering or audit skills are needed for site visit operable testing for some measures. The extent of the experience and expertise needed varies with the sophistication of the operational testing. Verification of make, model number, and likelihood that the piece of equipment is the original program installed one can be made by auditors or engineers with experience/training with regard to identification of the type of equipment being examined. Operable verification that uses commissioning reports, energy management system reports, or similar reporting must be reviewed by engineers with experience/training that allows a quality verification effort.

Telephone surveys and interviews need to be conducted by experienced personnel. These studies and their instruments must be designed with personnel with experience in energy efficiency markets, the social sciences, and interview and survey instrument design, implementation and analysis.

Technical degradation studies require senior experienced engineers that are quite familiar with the equipment to be studied, its standard counter-part, and the components, operations, and effects of changes in the operational conditions on the components and function of the equipment. Senior personnel must also have the time budgeted for significant input and review, for responsible project task leadership and quality control. The degree of involvement needed from senior skilled staff is dependent upon the skill and experience of the mid-level personnel conducting much of the analysis work.

Methods for conducting behavioral degradation could be based upon survey and interview analysis methods and/or statistical/econometric methods. The personnel to conduct the work need to have the skills and experience for the method being proposed. The evaluators need to be trained and experienced in conducting social science research with a strong understanding of assessing and testing causal relationships between exposure to the program and possible outcomes. An important requirement is for these evaluators to have a strong foundation in research design and the ability to create research designs to test for net behavioral impacts of energy efficiency programs.

Summary of Protocol-Driven Impact Evaluation Activities

1	Joint Staff will review retention, EUL, and degradation planning information, perhaps through an initial study of (1) prior retention, EUL, and degradation studies and methods, (2) required retention, EUL, and degradation sample sizes, and (3) assessment of data collection methods for the prioritized measure and delivery strategy/application needs. Along with any risk analysis information, Joint Staff will identify which measures by delivery strategy/application will receive which type of retention, EUL, and degradation evaluation, when, and at what rigor level. Joint Staff will determine any special needs on a case-by-case basis that will be required for particular retention, EUL, and degradation evaluations. Joint Staff will develop preliminary RFPs for groups of studies based upon timing of the needed data collection or analysis, similar sectors or issues to be addressed, and requiring similar skill sets. CPUC-ED will issue RFPs for retention, EUL, and degradation evaluations, select evaluation contractors, and establish scope(s) of work.
2	Evaluators will develop a research design and sampling plan to meet Protocol requirements as designated by the Joint Staff rigor level assignments. This includes meeting requirements from the Sampling and Uncertainty Protocol, as are applicable given Effective Useful Life Evaluation Protocol requirements. Research design and sampling must be designed to meet any of the Joint Staff requirements for additional analyses to include but not limited to areas designated of specific concern by the Joint Staff. Evaluators will develop and submit an Evaluation Plan to Joint Staff, and the plan will be revised as necessary to have an approved Evaluation Plan that meets the Effective Useful Life Evaluation Protocol.
3	All retention, EUL, and degradation study evaluation teams (including panel data collection teams) will make sure their teams are appropriately staffed, in order to meet the skills required for the research design, sampling, and selected retention, EUL, and degradation evaluation method, uncertainty analysis, and reporting being planned and conducted.
4	All retention, EUL, and degradation study evaluations will be planned, conducted, and analyzed to minimize potential bias in the estimates (showing the methods for doing this), and evaluators will report all analyses of potential bias issues as described in the Sampling and Uncertainty Protocol.
5	All retention, EUL, and degradation evaluations will be conducted according to the Evaluation Plan and appropriate Protocols.
6	Evaluators will develop the draft evaluation report in accordance to guidance provided by the Joint Staff and reporting requirements in this Protocol.
7	Final evaluation report will be developed in accordance to guidance provided by the Joint Staff, and then submitted to Joint Staff.
8	Once accepted by Joint Staff, abstracts will be developed, and a report will be posted on the CALMAC web site following the CALMAC posting instructions.

Note: the steps included in this evaluation summary table must comply with all the requirements within the Effective Useful Life Evaluation Protocol in order to be in compliance.

Process Evaluation Protocol

Introduction

The Process evaluation is not a required evaluation activity in California. It is, however, often critical to the successful implementation of cost-effective and cost-efficient energy efficiency programs. Process evaluations identify improvements or modifications to a group of programs, individual programs or program components, that directly or indirectly acquire or help acquire, energy savings in the short-term (resource acquisition programs) or the longer-term (education, information, advertising, promotion and market effects or market transformation efforts).

The primary purpose of the process evaluation is an in-depth investigation and assessment of one or more program-related characteristics in order to provide specific and highly detailed recommendations for program changes. Typically, recommendations are designed to affect one or more areas of the program's operational practices. Process evaluations are a significant undertaking designed to produce improved and more cost-effective programs.

This Protocol identifies how process evaluations for California energy efficiency programs, products or services placed into the market during and after the 2006 program year will be conducted. The *Evaluation Framework* is incorporated into this Protocol as a key guidance document for conducting process evaluations. Before applying it, users of this Protocol should have a working knowledge of Chapter 8, "Process Evaluations," of the *Evaluation Framework*.⁸⁸ Key references to the process evaluation literature are also found in the *Evaluation Framework*.⁸⁸

In addition, all users of the Process Evaluation Protocol should be familiar with the sampling guidance provided in the Sampling Protocol. While the use of the Sampling Protocol is not required for planning and conducting process evaluations, it provides guidance for process evaluation sampling and sample selection criteria. The Reporting Protocol contained in this document contains the process evaluation reporting requirements and is a part of the Process Evaluation Protocol. The Reporting Protocol helps assure that the reports generated from the process evaluation provide comparable results and recommendations on which program management can act, and, at the same time, meet the CPUC's reporting requirements.

A process evaluation is defined as a systematic assessment of an energy efficiency program, product or service, or a component of an energy efficiency program, product or service, for the purposes of identifying and recommending improvements that can be made to the program to increase the its efficiency or effectiveness in acquiring energy resources while maintaining high levels of participant satisfaction and documenting program operations at the time of the examination. The primary goal of the process evaluation is the development that improve program efficiency or effectiveness that when implemented can be expected, in some direct or indirect way, to improve the cost effectiveness of the program. This definition updates the definition provided in the *Evaluation Framework*.⁸⁹

⁸⁸ TecMarket Works, page 205.

⁸⁹ Ibid, page 207.

Process Evaluation Responsibilities

While the CPUC- ED must approve all process evaluation contractors, the process evaluations themselves are to be planned, budgeted, designed, implemented and reported under the direction of the Administrators, following the guidance laid out in this Protocol. The Administrators are responsible for the process evaluations for their statewide programs, the Administrator-specific programs implemented within their services areas, the programs conducted by third parties under contract to the Administrators and the programs or services that are procured via a bidding or other contractual processes. The Administrators are responsible for developing the Annual Process Evaluation Plan and obtaining related comments and recommendations from the Joint Staff.

Objectives of the Process Evaluation

The process evaluation's primary objective is to help program designers and managers structure their programs to achieve cost-effective savings while maintaining high levels of customer satisfaction. The process evaluation helps accomplish this goal by providing recommendations for changing the program's structure, management, administration, design, delivery, operations or targets. Consequently, Administrators often want early process evaluation feedback. The process evaluation also provides ongoing feedback to the Administrators that allow them to make timely program changes or to follow the progress of the study or to review early findings. Where appropriate, the process evaluation should test for the use of best practices and determine if specific best practices should be incorporated. It is expected that process evaluations will be needed both in the early stages of the program design and deployment efforts to provide timely feedback to the IOUs on them, and over the life of the program as issues are identified.

Audience and Responsible Actors

This Protocol is to be used by Administrators and their evaluation contractors to conduct process evaluations and by the CPUC-ED to provide ongoing guidance and oversight to the process evaluation activities. The Protocol allows considerable flexibility and judgment by the Administrators to determine when a process evaluation is needed and the issues on which the process evaluation should focus.

The Administrators are responsible for program design, operation and goal attainment for the programs and services funded through their implementation and contracting efforts. They must structure their process evaluation efforts to support these responsibilities.

Other stakeholders should be familiar with the intent and scope of the Protocol to obtain an adequate understanding of the purpose and scope of the process evaluation efforts and how these studies are to be conducted.

Overview of the Protocol

As mentioned previously, this Protocol is specifically designed to work in conjunction with the *Evaluation Framework*. The chapter on process evaluation in the *Evaluation Framework* is a key advisory component of this Protocol.⁹⁰

⁹⁰ Ibid, page 205.

This Protocol provides guidance to Administrators on the criteria used to determine if and when to conduct a process evaluation and on the researchable issues targeted by the study.⁹¹ However, Administrators are free to identify additional or different decision criteria.

Because a process evaluation is not a CPUC-required activity, the Administrators will determine if one is to be conducted, when it will occur and the investigative areas on which the evaluation shall focus. As a result, there is no waiver process associated with the Process Evaluation Protocol. However, the Protocols suggest several key investigative areas on which the process evaluation can focus.

Finally, the Protocol presents the types of investigative tools and approaches that can be used to conduct the process evaluation efforts.

As discussed earlier, the Process Evaluation Protocol is linked with other of the Protocols, including the Sampling and Uncertainty Protocol and the Reporting Protocol. These two latter Protocols should be considered sub-components of the Process Evaluation Protocol. However, the Sampling and Uncertainty Protocol is advisory, while the Reporting Protocol is a required component of the Process Evaluation Protocol.

Because there can be overlap between the information collected in the process, market effects and impact evaluations and associated M&V efforts, the process evaluation efforts should be structured to coordinate with other evaluation efforts to the extent practical. This may require cross-organizational coordination. Such coordination minimizes customer contact, maximizes data collection efforts and improves evaluation efficiency. This does not mean that these studies must be inter-linked or consolidated, but it does mean that there will be times when the information collected in one study will be valuable to other studies and times when studies will benefit from a coordinated effort.

In these cases, there will need to be close coordination between the CPUC-ED, the Administrators and the evaluation contractors for the related evaluation efforts. This Protocol does not specify how this coordination should be structured or conducted, but does identify the need for it. However, the Protocol also recognizes that the skill sets needed for conducting a process evaluation are often different than those needed for an impact or market effects study and these differences may limit the extent of the coordination efforts.

Process Evaluation Planning

There are several key issues that should be considered when planning a process evaluation. It is anticipated that most programs will have at least one in-depth comprehensive process evaluation within each program funding cycle (e.g., 2006-2008), but a program may have more or less studies depending on the issues that the IOUs need to research, the timing of the information needed and the importance of those issues within the program cycle.

The process evaluation decision road map in the *Evaluation Framework*⁹² identifies several operational conditions that can be considered by the IOUs for targeting a process evaluation.

⁹¹ Ibid, page 220.

The Administrators should assess these criteria annually and other criteria as appropriate to determine if a process evaluation is needed. This annual assessment should be conducted by the IOUs administering the program and reported in an Annual Process Evaluation Plan delivered to the CPUC-ED no later than the first of December, before the start of each program year or the month before the start of a new program when the program does not start at the beginning of a program year.

This annual planning requirement applies to all programs being administered or funded via the Public Goods Charge (PGC) or Procurement program funds for the upcoming program year. While the detailed process planning efforts will not be fully known at the beginning of the program cycle, the plan should present the structure and operations of the detailed planning efforts, identify the key decision criteria to be used to determine if and when a process evaluation will be planned and launched, and present the process evaluation efforts planned at that time. The annual plans developed following the first year would then present the process evaluations planned and launched during the previous year and present the known process evaluation needs for the upcoming year.

Annual Process Evaluation Planning Meeting

For each year of the program cycle, the Administrators shall hold at least one process evaluation planning meeting to review and discuss their process evaluation plans and obtain Joint Staff input to help guide planning efforts. The meeting shall be held between July and November of the year preceding the evaluation period. The Joint Staff will be notified of these meetings at least two weeks in advance of the meeting. The meeting dates, times and locations will be coordinated with the Joint Staff to maximize the attendance potential of the interested parties. During the meeting, the Administrators will present their process evaluation plans and solicit comments. Within two weeks following the meeting, but no later than December 22, the Administrators will finalize their process evaluation plans and submit a final plan to the Joint Staff for its review. The plan does not have to be approved by the CPUC-ED.

The decision for determining if a process evaluation is needed for all programs rests with the Administrators. However, the evaluation planning process must be conducted annually, incorporate the use of the planning meetings discussed above and lead to the submission of an Annual Portfolio Process Evaluation Plan to the Joint Staff for review.

Program managers may want the process evaluation to supplement the program's quality control or quality assurance components, to confirm the installation practice and/or to conduct program reviews and develop recommendations for improvements. Administrators should consider the different functions and benefits of the process evaluation under different potential program grouping scenarios and weigh the associated pros and cons when structuring their process evaluation plans.

The timing of the process evaluation is an important component of the planning process. In some cases, Administrators will want to launch their process evaluations early to help ensure that programs are well designed, are achieving savings shortly after launch and are providing

⁹² Ibid, page 222.

effective services. In other cases, Administrators may give the programs time to become established in the market and adopt more routine operational approaches before launching the process evaluation. In still other cases, Administrators may wish to establish an ongoing process evaluation effort so that the program is periodically evaluated over its three-year cycle. This third condition may be more applicable to new programs or programs being provided by a new vendor than to more established programs that have demonstrated their cost-effectiveness and operational efficiencies in earlier studies.

The Annual Process Evaluation Plan submitted to the Joint Staff should indicate the level of resources and budgets devoted to the process evaluation efforts.

The goal of this Protocol is not to require unnecessary process evaluation efforts, but to provide tools to Administrators for the consideration of key decision criteria typically associated with process evaluation efforts and related implementation. Administrators should only plan and launch process evaluations when they are expected to serve as an effective program management tool. The Annual Process Evaluation Plan helps to provide the Joint Staff with a minimum level of assurance that program changes are being effectively assessed and managed and that process evaluations are considered when they can be effectively employed. The decision to conduct a process evaluation is ultimately the Administrators'. These decisions should be conveyed to the CPUC-Ed in the Annual Process Evaluation Plan.

Recommendations for Change

The primary purpose of process evaluation is to develop recommendations for program design or operation changes that can be expected to cost-effectively improve the issues, conditions or problems being investigated. The primary deliverable of all process evaluations is a process evaluation report that presents the study findings and the associated recommendations for program changes (see Reporting Protocol).

Key Issues and Information Covered

Administrators and their need for operational information to improve programs guide the process evaluation effort. This necessarily covers a very wide range of investigative issues that the process evaluation can address. The process evaluation may take on the challenge of evaluating most, if not all, aspects associated with the design or operations of a program in order to improve the energy resources acquired (directly or indirectly) by that program. The process evaluation plan can also address issues applicable to the programs under review over a single year or over multiple years and can examine a wide range of issues, including:

Program Design

- Program design, design characteristics and design process;
- Program mission, vision and goal setting and its process;
- Assessment or development of program and market operations theories and supportive logic models, theory assumptions and key theory relationships - especially their causal relationships; and
- Use of new practices or best practices.

Program Administration

- Program oversight and improvement process;
- Program staffing allocation and requirements;
- Management and staff skill and training needs;
- Program information and information support systems; and
- Reporting and the relationship between effective tracking and management, including both operational and financial management.

Program Implementation and Delivery

- Description and assessment of the program implementation and delivery process;
- Quality control methods and operational issues;
- Program management and management's operational practices;
- Program delivery systems, components and implementation practices;
- Program targeting, marketing and outreach efforts;
- Program goal attainment and goal-associated implementation processes and results;
- Program timing, timelines and time-sensitive accomplishments; and
- Quality control procedures and processes.

Market Response

- Customer interaction and satisfaction (both overall satisfaction and satisfaction with key program components and including satisfaction with key customer-product-provider relationships and support services);
- Customer or participant energy efficiency or load reduction needs and the ability of the program to provide for those needs;
- Market allies interaction and satisfaction;
- Low participation rates or associated energy savings;
- Market allies needs and the ability of the program to provide for those needs;
- Reasons for overly high free-riders or too low a level of market effects, free-drivers or spillover; and
- Intended or unanticipated market effects.

Process Evaluation Efforts

One of the primary purposes of the Process Evaluation Protocol is to ensure an appropriately broad consideration of potential process evaluation issues for each program within the Administrator portfolio and a framework that provides critical evaluation thinking to produce the best overall process evaluation efforts for the portfolio.

Program-Specific Process Evaluation Plans

In addition to the Administrator's Annual Portfolio Process Evaluation Plan, each process evaluation should have a program-specific or program-group-specific detailed process evaluation plan to guide the evaluation efforts. These detailed plans should include the process evaluation approach, identification of program-specific or program group-specific focus of the evaluation efforts, detailed researchable issues to be addressed, activity timing issues and the resources to be used. However, it is the Administrator's responsibility to specifically determine the content and focus of such plans. The detailed program-specific or program-group-specific detailed process evaluation plans do not need to be submitted to the Joint Staff for review.

The Process Evaluation Protocol is designed to balance allowing the CPUC-ED and other stakeholders a level of assurance that there is a minimum set of standards for process evaluations across the portfolios and allowing the necessary flexibility and control for program administration and process evaluation management.

Data Collection and Assessment Efforts

Process evaluation efforts can include a wide range of data collection and assessment efforts, such as:

- Interviews and surveys with Administrators, designers, managers and implementation staff (including contractors, sub-contractors and field staff);
- Interviews and surveys with trade allies, contractors, suppliers, manufacturers and other market actors and stakeholders;
- Interviews and surveys with participants and non-participants;
- Interviews and surveys with technology users;
- Interviews and surveys with key policy makers and public goods charge stakeholders;
- Observations of operations and field efforts, including field tests and investigative efforts;
- Unannounced participation in the program to test operations and operational practices, processes and interactions;
- Operational observations and field-testing, including process related measurement and verification efforts. These can be announced or unannounced;
- Workflow, production and productivity measurements;
- Reviews, assessments and testing of records, databases, program-related materials and tools used;
- Collection and analysis of relevant data or databases from third-party sources (e.g., equipment vendors, trade allies and stakeholders and market data suppliers); and
- Focus groups with participants, non-participants, trade allies and other key market actors associated with the program or the market in which the program operates.

This list of activities is not meant to be exhaustive, but illustrative. Administrators are free to specify other data collection and assessment efforts beyond those listed above. However, in selecting the evaluation approaches to be used, a key consideration is the level of reliability of the study approach and the accuracy of the study findings. All studies are expected to be structured in a way that provides reliable findings on which accurate and reliable recommendations can be developed.

Conducting Investigative Efforts

This section of the Protocol provides guidance on conducting specific investigative efforts typically associated with the process evaluation.

Interviews

Professional process evaluation interviewers should conduct process evaluation interviews. The *Evaluation Framework* provides guidance on the type of training and experience needed by process evaluation staff that conduct interviews.⁹³ In-depth interviews can be conducted in-person (off-site or on-site) or by telephone. In-person interviews enable the interviewer to gain a deeper understanding of the experience of the interviewee and can lead to more reliable and more comprehensive information gathering than phone interviews. Phone interviews do not allow for the observation of key body signals that serve as clues for the probing process. However, both approaches are equally valid if the questions are well designed and the interviewer is skilled in interviewing techniques. If in-person interviews are not possible given the nature of the study or the location of the interviewee, then telephone interviews can be used. Regardless of the type, interviews should be scheduled in advance and should last an hour or more. Detailed comprehensive process evaluation interviews may last several hours. E-mail interviews are rarely used unless the evaluation professional can easily guide the interview and move it in directions that need additional information or investigative probes. In addition, as technologies that can be used to support the interview effort evolve, there may be additional approaches that can be considered or used, such as web-conferencing or web-interviews. However, in assessing the applicability of these technologies the primary focus should be on allowing the interviewer to be able to manage and focus the interview as it proceeds so that in-depth probes and ancillary follow-up questions can be placed into the interview at the time they are needed.

Group Interviews

The group interview may be used to obtain information from a group of individuals typically having one or more similar characteristics. The focus group, one of the more familiar types of group interviewing techniques, is used to focus on the response to a limited set of issues – such as in product development research. The use of other types of group interviews can be appropriate for evaluation in a limited number of circumstances, for example:

- Obtaining feedback from a group of installers or outreach coordinators who can “focus” on the specific issues of their job or their experience with end-users; and
- Qualitatively investigating issues that will be further explored in quantitative surveys.

Experienced professional evaluation experts must conduct group interviews. All group interviews should be guided by skilled moderators and documented in a way that allows for a

⁹³ Ibid, page 206.

review of the moderator's instructions to the attendees, the moderator's approach to managing the group and the moderator's instructions, questions and involvement, as well as a detailed understanding of comments provided by all attendees.

A summary report of each group interview containing the above-listed information should be included in the evaluation report. The group interview report should also include a professional interpretation of the results discussing how they will or can be used and a discussion of how the group results will be confirmed or tested quantitatively, if required.

Group interviews are a reliable data collection approach but they do not provide results that can be generalized to a population except when the participants in the group interview are a statistically representative sample of that population. Thus, a focus group (because the participants do not typically constitute a statistically representative sample of the population) is not an acceptable means to quantitatively assess programs, but can, in some circumstances, provide supportive information that can be used in a process evaluation finding. For this reason, they are included as an approved data collection effort within this Protocol when accompanied by other assessment approaches that can quantitatively test their results.

Surveys

Survey efforts that are used to support process evaluations are typically conducted via telephone interviews. However, there will be occasions when other approaches are preferred. For instance, when there are large numbers of participants or non-participants or when the inquiry will benefit from the respondent seeing an illustration, survey techniques could include mail, e-mail or Web-based approaches and other types of surveys. Similarly, small targeted surveys with trade allies or program participants who have provided e-mail addresses for this purpose may be most efficiently conducted using an e-mail/Internet combined survey. There is a great variety of survey techniques and they should be selected according to specific requirements of the data collection effort.

In all cases, professional process evaluation survey designers should construct and test the survey questionnaires to avoid unnecessary bias in question topic or structure, or in the responses received. The questions in the surveys should follow construction practices that result in objectively worded questions with provisions for recording all expected responses. Questions should be structured so that they are single-subject, focused questions. Questions that are typically referred to as "double-barreled" questions (containing more than one subject-verb relationship) should be avoided as they bias the information collected.

Most important in implementing any type of survey is to follow the principles of good survey design and implementation such as those developed by Don Dillman.⁹⁴ Whether the survey is implemented using the telephone, mail, e-mail or Internet there are specific methods that should be applied to ensure valid and reliable results. These include repeated contacts with the sample as well as carefully structured invitations to participate and questions.

⁹⁴ Dillman, Don A. 2000. *Mail and Internet Surveys: The Tailored Design Method*. New York, NY: John Wiley & Sons.
Dillman, Don A. 1978. *Mail and telephone surveys: The Total Design Method*. New York, NY: John Wiley & Sons.

Observations and Field Testing

Field-testing and observations should be done in a way that allows the observation or testing of the program as it would be operating in the absence of the evaluation professional. Observers are to instruct program staff that they are to conduct themselves exactly as they would if the observer were not present. The observing evaluation professional is not to engage in activities that act to change the way the activity would have occurred if the evaluation professional were not there. All key observations and measurements should be documented at the time of the observation or testing.

Unannounced Participation

In some cases, it may be appropriate for the evaluation contractor to enroll in the program to test the program's operations and delivery aspects. When this is designed as part of the evaluation plan, program management is not to be informed of who will be participating, how they will be participating or when that participation will occur. Participation is to be unannounced and field observations and measurements will be conducted without the knowledge of the program staff to the extent practical. This approach can be used for a wide range of programs in which unannounced participation by an evaluation professional can allow the evaluation expert to view the program from the perspective of a typical participant.

Independence

The organization conducting the process evaluation should be independent of the organizations involved in the program design, management and implementation efforts. The evaluation should be conducted as an "arms-length" assessment, such that the process evaluation professionals have no financial stake in the programs or program components being evaluated beyond the contracted evaluation efforts. Similarly, process evaluation professionals should have no financial or financially related interest in the study results or from efforts resulting from the implementation of evaluation recommendations.

Selection of Evaluation Contractors

Administrators are charged with the responsibility to plan, contract, manage and administratively oversee the implementation of the process evaluation efforts consistent with this Protocol. Administrators should focus their contractor selection efforts, so that only professional, skilled process evaluation contractors are solicited for conducting the process evaluations. The CPUC-ED must approve the selection of the evaluation contractors to conduct the studies. The contractor approval process will be structured by the CPUC-ED consistent with the ALJ's decision. This process will be developed outside of this Protocol. Approval by the CPUC-ED will be based on the qualifications of the firms or individuals considered for conducting the studies.

Skills Required for Conducting Process Evaluations

The investigative processes associated with designing, managing and conducting process evaluations focus on a wide range of researchable issues. These issues can range from evaluating the ability of a program's data management system to support the informational needs of the program to assessing if the program is well-designed, managed, targeted, marketed and operated. As a result, the skills needed to conduct process evaluations are varied.

Evaluations that focus on the design and operation of program information systems, for example, need evaluators that understand how information management and information availability influence a program's management, operations, productivity and results. However, the evaluators should also be skilled at designing, developing and implementing information systems in order to recommend changes to improve the program's ability to cost-effectively achieve its goals. Process evaluators who assess program satisfaction levels need to have the skills to identify and analyze different program characteristics that influence satisfaction and be able to identify those characteristics that can be changed to improve satisfaction scores. In the process evaluation, measuring satisfaction is not enough, the study should assess the reasons for the satisfaction scores and identify how to improve these scores without harming the cost-effectiveness of the program.

Similarly, evaluators who focus on assessing program targeting, marketing and promotional operations need to have skills necessary to assess information flow, content and presentation effects as well as the skills associated with understanding how markets and market segments operate and can be influenced through different outreach and promotional efforts. These examples demonstrate the need to match the skills of the process evaluator with the research goals of the specific process evaluation.

It is equally important that process evaluation managers be trained and/or experienced with the tools used in the process evaluation. For example, if a telephone survey is needed, evaluators need to be knowledgeable and experienced in the field of survey research and instrument design. If focus groups are needed, evaluators should be knowledgeable and experienced in the field of focus group design and operation, as well as assessing and applying the results from the focus group.

Because of the diversity of researchable issues associated with conducting process evaluations and the diversity of skills needed to address these issues, it is difficult to define a specific set of skills needed to conduct these evaluations. Instead, this Protocol recognizes that a diverse set of program assessment and information analysis skills are needed across the various investigative issues on which these evaluations typically focus. However, in general, the process evaluator should have the following knowledge and skills:

- Expert knowledge of a wide range of energy efficiency programs and a strong understanding of their operational designs, management practices and program goals;
- Expert knowledge of different process evaluation data collection methods and approaches, and a working knowledge of the process evaluation literature and how evaluation approaches have been applied in the energy efficiency program field;
- Strong analysis capabilities and an expert understanding of cause-and-effect relationships that impact the ability of energy efficiency programs to cost-effectively accomplish their goals, including experience in program theory and logic model construction and assessment;
- Strong understanding of statistical analysis approaches and analytical procedures appropriate for the process evaluation research goals;

- Strong understanding of sampling methods and approaches and the ability to identify potential biases in a sampling approach and to develop control strategies for mitigating levels of bias and improving the reliability of evaluation results; and
- High level of past experience in conducting process evaluations of energy efficiency programs and in reporting the results of these studies.

Market Effects Evaluation Protocol

Introduction

The Market Effects Protocol is designed to measure net market effects at a market level when one or more of the Protocol-covered energy efficiency funded program efforts target a market. Net market effects are those effects that are induced by Protocol-covered energy efficiency programs and are net of market activities induced by non-energy efficiency programs including normal market changes.

The application of the Market Effects Protocol should result in an estimate of the energy (kWh), peak (kW) or therm impacts associated with the net market effects resulting from Protocol-covered energy efficiency program interventions. These net energy market effects are referred to in *A Framework for Planning and Assessing Publicly Funded Energy Efficiency (2001 Framework Study)* as “ultimate market effects” or “ultimate indicators” because they are the desired indicator of whether net energy efficiency changes are occurring in the market.⁹⁵ The Market Effects Protocol is designed, therefore, to facilitate not just the estimate of net market effects but also, and primarily, the estimate of net energy market effects. That is, a market effects study both quantifies the changes occurring in the market caused by the energy efficiency programs and provides an estimate of the energy impacts associated with them.

The Market Effects Protocol does not apply to the measurement of individual program-level market effects or direct program savings typically used for program-level cost-effectiveness assessments and refinement decisions. Rather the focus of the market effects evaluation is at a market level in which may different energy efficiency programs can operate. Yet, the Protocol applies to program-induced market changes that could be missed or double counted if measured program by program. As a result, the use of the Market Effects Protocol should focus on the effects of groups of programs within a market over multiple program cycles.

Overview of the Market Effects Protocol

This Protocol applies when net market effects are to be estimated at a market rather than program level. Market effects are defined in the *Evaluation Framework* as “[a] change in the structure or functioning of a market or the behavior of participants in a market that result from one or more program efforts. Typically these efforts are designed to increase the adoption of energy efficient products, services, or practice and are causally related to market interventions.”⁹⁶ This definition, however, was created within the context of guidance for conducting program evaluation of a market transformation style program. A market transformation program is one that is specifically designed and fielded for the purpose of changing the way a market operates so that energy savings are achieved at a market level. That is, these types of programs are designed to focus at the market level. A more effective definition for the Market Effects Protocol for assessing the market effects from multiple programs that may or may not be designed to change market operations is that in *A Scoping Study on Energy-Efficiency Market Transformation by*

⁹⁵ Frederick D. Sebold et al. *A Framework for Planning and Assessing Publicly Funded Energy Efficiency*. (Pacific Gas and Electric Company. 2001): 6-4.

⁹⁶ TecMarket Works, 429.

California Utility DSM Programs (the Scoping Study): “A change in the structure of a market or the behavior of participants in a market that is reflective of an increase in the adoption of energy-efficient products, services, or practices and is causally related to market intervention(s).”⁹⁷ This definition stresses the market rather than the program nature of market effects, and is the working definition for this Protocol.

The *Evaluation Framework* states that “there are no universally accepted definitions within the energy efficiency industry pertaining to what constitutes a program’s market.”⁹⁸ A review of various dictionaries demonstrates that it has multiple meanings. “Market” as used in this Protocol refers to the commercial activity (manufacturing, distributing, buying and selling) associated with products and services that affect energy usage. The specific market focus of each evaluation should be defined as an early activity in scoping each market effects evaluation. The *Evaluation Framework* provides guidance for defining a market and where multiple programs are operating in the same market, again, the primary focus of this Protocol.⁹⁹

Market effects include both short-term and long-term effects. The long-term effects are the most difficult to capture at a program level because they broadly affect a market not just the specific participants in a program or in a grouping of programs. This Protocol targets those long-term effects.

A market-level evaluation effort is recommended when there are multiple statewide or local interventions in a market such as those of California’s energy efficiency programs and where other efforts are also acting to change that market. Other efforts can be associated with the normal operations of the market or when other non-California energy efficiency efforts are changing markets, such as with the national ENERGY STAR[®] program, manufacturer promotions and retail sales efforts. A market level effort is also appropriate when a single large and particularly effective program is expected to have broad and long-term market effects in a single market.

Figure 11 shows the relationship between program-induced market effects, program market spillover and normal energy efficiency trends in the market. Effects driven by interventions by other organization, as well as the market itself, are all assumed to be within the normal energy efficiency trends of the market.

⁹⁷ Joe Eto, Ralph Prael, and Jeff Schlegel. *A Scoping Study on Energy-Efficiency Market Transformation by California Utility DSM Programs*. (Berkeley, CA: Ernest Orlando Lawrence Berkeley National Laboratory, 1996). LBNL-39059 UC-1322, 9.

⁹⁸ TecMarket Works, 250.

⁹⁹ Ibid, 250-251.

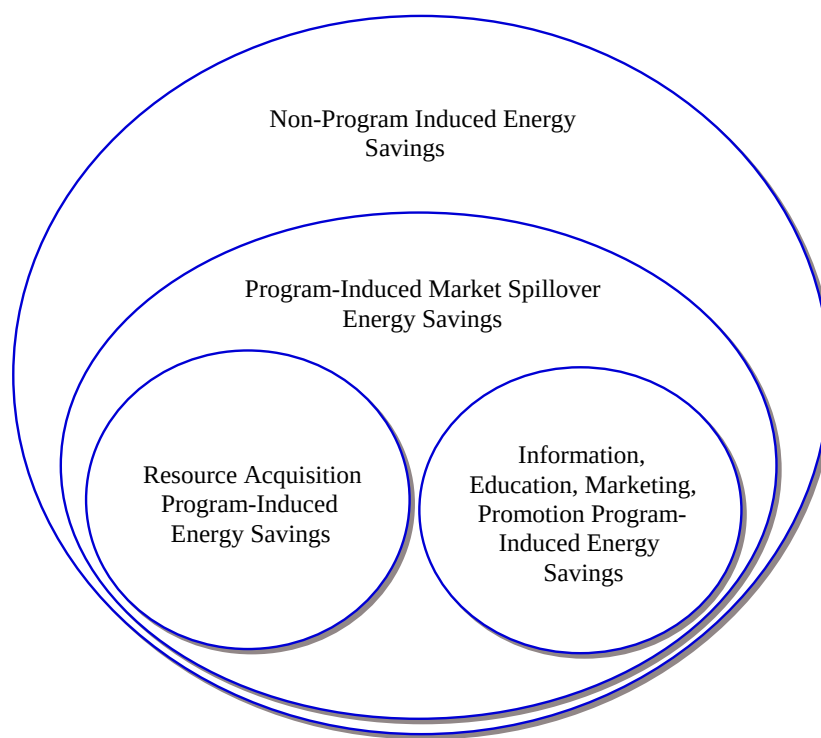


Figure 11. Sources of Energy Efficiency Changes in the Market

There are two types of market effects discussed in the energy efficiency industry. There are those that are occurring now as a result of how programs are changing markets. And there are those that are forecasted to occur later (after the program has been discontinued) due to the changes established or put into motion by the program. The Protocol recognizes that the methodologies to estimate each of these types of market effects can differ and that potential issues of bias that must be identified, mitigated and minimized are also different. The Market Effect Protocol is designed to measure only the current market effects and not those forecasted to occur at some future point.

A great deal of effort has been expended over the past 10-15 years to estimate market effects, yet most of these efforts did not estimate net energy market effects, but concentrated on measurement of indicators such as awareness, sales and changes in practices by market actors. Evaluations estimating net market effects with energy estimates, the focus of this Protocol, are at an early stage of development. A variety of studies have been conducted, but only a limited number at the highest levels of rigor. However, this is a critically important field of research since the market effects of energy savings caused by California's energy efficiency programs are likely to be substantial once documented. Given the early stage of development of methods, it is important that this Protocol encourage the continued advancement of the field and not prescribe or limit methodological approaches.

Key to a successful market effects evaluation will be the initial scoping study. The scoping study will define the market to be studied, develop a market theory to test in the analysis, assess data availability for the market effects study, develop a methodology for additional data collection and recommend an analysis approach. For programs that are specifically designed to change the way a market operates, the program theory should also be considered in developing the initial scoping study. However, for standard programs that are not designed to change market operations, the program theory is not a significant consideration in the development of the scoping study.

Because market effects evaluation is still evolving there are a limited, but clearly defined, set of activities that should be considered. Market effects evaluations should be developed using experimental or quasi-experimental designs whenever possible and the approach should be peer reviewed prior to implementing the study to ensure that it will provide valid and reliable results. Triangulation of data and analysis approaches is preferred when possible and teaming with industry organizations and professionals can be beneficial. The studies should also take into account regional differences within the market being studied and will at times need to move beyond California boundaries to the regional or national level to collect data. Finally, allocation to utility service territory will be a challenge and dependent on data availability, but should be an important consideration in the scoping study.

The Market Effects Protocol and Other Protocols

Often the individual Protocols overlap and are supported by other Protocols. There are three primary output Protocols: Impact, Market Effects and Process Evaluation. There are three Protocols that can be called on to support or provide additional requirements for all the Protocols. These are the Sampling and Uncertainty, Measurement & Verification (M&V) and Reporting Protocols. The guidance provided by the Impact Evaluation, Market Effects and Process Evaluation Reporting Protocols applies to all types of efficiency program evaluations in California. The Sampling and Uncertainty Protocol provides further delineation of sampling and uncertainty assessment and reporting requirements. The M&V Protocol is similarly a reference and supporting Protocol for the Impact Evaluation Protocol and can be used to inform and provide input for a Process Evaluation on issues relating to measure installation and performance.

The Impact Evaluation and M&V Protocols are supporting Protocols to the Market Effects Protocol as related to estimating net energy market effects. At the same time while the Impact Evaluation Protocol addresses net energy effects through estimation of free-ridership and participant spillover, it does not include measurement of non-participant spillover. Non-participant spillover specifically refers to changes in the market that result from program influences and this is appropriately estimated as part of the net market effects.

There is another important integration aspect for the individual Protocols. A complete measurement of program impacts involves combining the results from the market effects evaluations and the program impact and indirect impact studies. The market effects net of program impacts would generally represent the program market level impacts and non-participant spillover. Yet, differences in methodologies and the multitude of program

evaluations require careful thought and analysis to ensure that when the impacts are combined adjustments are made to ensure no double counting occurs. At the same time, there are bound to be some programs for which program impacts have been evaluated but their particular market has not (e.g., in cases of niche markets or unique program elements). This would need to be added to reach an estimate of the overall portfolio expected energy and demand savings to be reported to system planners. Familiarity with the Impact Evaluation, M&V, Sampling and Reporting Protocols is recommended to conduct a Protocol-compliant Market Effects Evaluation.

Finally, the Process Evaluation Protocol outlines types of data collection methodologies that should also be considered when conducting primary data collection for estimating market effects. The *Evaluation Framework* provides further detail on data collection methodologies and issues to consider when examining markets, familiarity with which is recommended for designing or conducting a Market Effects Evaluation.¹⁰⁰

Key Market Effects Inputs and Outputs

Inputs for a market effects evaluation include but are not limited to the following:

- Names and contact information for program staff for the programs identified as targeting the defined market;
- Names and contact information for mid-stream or upstream market actors identified by the Administrator as operating in the defined market;
- Evaluations and market research conducted by the utilities for the defined market during the previous five years;
- Market and program theory documents developed by the Administrators for the programs identified as targeting the defined market;
- Names and contact information for key informants consulted during the development of the programs identified as targeting the defined market; and
- National data on the market from sources such as the US Census Bureau, Energy Information Administration, Consortium for Energy Efficiency, Department of Energy and Environmental Protection Agency.

The Market Effects Protocol will generate an estimate of net energy market effects in kWh and kW in markets for electricity-using equipment and therms in markets for gas-using equipment. These metrics at times will require that the market effects estimates link to the results of the M&V or Impact Evaluation Protocol to provide estimates of energy impacts based on the market effects measured. The market to which these estimates apply will be defined in terms of location, the utilities involved, the equipment and sector, and the program years of interest.

This approach requires that the evaluator estimate what changes would have occurred in the market without the energy efficiency efforts provided by the programs. Because of the uncertainty in the evaluation process, the estimate will likely be a range of probable effects, rather than a point estimate (e.g., confidence intervals). These studies should always include a clear statement of the uncertainty around the range estimate. The Reporting Protocol discusses this issue in greater detail.

¹⁰⁰ TecMarket Works, Chapters 4, 8, 9, 10, 12, and 13.

The outputs will be used to inform the program planning process for the next program cycle, to structure the program planning efforts to maximize net market effects of statewide and Administrator's portfolios, while also maximizing net program-induced energy impacts. The market effects evaluations should be structured to provide market effects impact information prior to June before the end of the program cycle so that the results can be considered in the planning efforts for the next program cycle.

The market effects study results can also be used in comparison with the results of the program evaluation efforts to identify net market energy impacts that are beyond the direct program-induced effects. In addition, results will serve as an estimation range check for the savings projected from the program evaluations. In all cases the total market effects should be a summation of the direct program-induced effects, the normal market changes to become more energy efficient and the non-participant spillover effects. In no cases should the program-induced impacts be greater than the net market effects identified in the market effects evaluations. Given the different methodologies employed, the analysis of the different sets of results, the methodologies used, and their relative advantages and weaknesses must be carefully conducted in order to obtain the most reliable estimates of overall impacts from the energy efficiency programs.

Audience and Responsible Actors

The audience and responsible actors for this Protocol include the following:

- **CPUC-ED and CEC** will use the Protocol to determine when a market effects study is appropriate and to guide the research approach;
- **The Evaluation Contractor Team** will use the Protocol to ensure that their market effects evaluation plan and its conduct address key requirements for a market effects study;
- **Administrators** will use the Protocol to understand the market interactions that are occurring as a result of efforts within a given market and, in part, to determine when interventions should be modified to achieve continued efficiency gains; and
- **Program Implementers** will use the results to assess the reach and success of their program efforts and in part to determine when interventions should be modified to achieve continued efficiency gains.

Steps in Conducting Market Effects Evaluations

The following five primary activities should be conducted in any market effects evaluation. This section describes in some detail what is entailed in each step.

1. Conduct a scoping study to determine optimum data collection and analysis approach for the evaluation;
2. Select a contractor and develop a detailed evaluation plan;
3. Collect baseline and longitudinal indicators;
4. Analyze market effects; and

5. Produce the Market Effects Report.

Scoping Study

The appropriate approach for a market effects study cannot be readily determined without a scoping study to define the market to be studied, develop a market theory to test in the analysis, assess data availability for the market effects study, specify a model of market change, develop a methodology for data collection and recommend an analysis approach.

Scoping studies will require different levels of effort depending on the complexity of the market of interest and the number and types of program interventions in that market. Scoping studies can also be used to determine which markets show promise for reliable and valid market effects evaluation.

The evaluation contractor will be expected to review past studies conducted of the market by the California utilities and other energy organizations. It will not be enough to simply look at the programs being offered during the program years of interest. A thorough review of the CALMAC database for applicable studies and reports as well as interviews with contacts at each utility and program managers will be important. Access to market assessment studies conducted by the utilities will be important to provide a sound understanding of the market conditions prior to program implementation and to support an understanding of market progress and/or contribute to the preponderance of evidence for causality/attribution. The evaluation contractor should also review the potential value of national and regional data sets including data collected by the US Census Bureau and the Energy Information Administration and organizations such as the Consortium for Energy Efficiency. No potential source should be ignored and it is very important that the utilities be cooperative in this effort.

The key activities of the scoping study include defining the scope range and limits of the subject market. The Joint Staff will make an initial determination of the definition of the market by location, utilities involved, sector, and likely equipment and behavior to be included. The scoping study contractor, however, will assess this initial definition and ultimately determine the definition that will yield the most reliable and meaningful results about net market effects of interest. This is a critical first step and requires a full understanding of how the interventions of interest were designed to operate and how the market in which they were launched is perceived to operate.

This process will provide the framework for the development of the market theory and conducting a logic analysis of the interventions, which will be used to guide the market effects evaluation. The logic model and market theory will then be used to develop a list of indicators for tracking market effects. These indicators could be model specifications, a detailed list of indicators to be tracked through baseline and longitudinal data collection efforts or both. The end result of the scoping study is an evaluation plan that details the strategy for the market effects evaluation. Table 14 displays the Protocol for scoping studies for market effects evaluations.

Table 14. Required Protocols for Market Effects Evaluation Scoping Studies

Level of Rigor	Scoping Study Requirements
Basic	Define the market by its location, the utilities involved, the equipment, behaviors, sector and the program years of interest. Develop market theory. Identify available secondary data and potential sources for primary data. Outline data collection and analysis approaches
Enhanced	Define the market by its location, the utilities involved, the equipment, behaviors, sector and the program years of interest. Develop market theory and logic model. Detail indicators. Identify available secondary data and primary data that can be used to track changes in indicators. Outline data collection approach. Recommend hypotheses to test in the market effects study. Recommend the analysis approach most likely to be effective.

Market Theory and Logic Models

The assessment, refinement and/or development of a market theory with logic models are key activities of the scoping study. The *2001 Framework Study*¹⁰¹ and the *Evaluation Framework*¹⁰² both address the value and process of developing a program or market theory. The evaluation contractor will need to articulate a market theory in order to proceed with baseline measurement for market effects evaluation. At a minimum, this market theory shall describe how the market operates and articulate market assumptions and associated research questions. This must be done at a level of detail sufficient to develop data collection instruments for baseline measurement. If the assessment includes programs that are designed specifically to change the way a market operates the program theory should also be consistent with and embedded in the theory of how the market operates.¹⁰³

Market-level evaluations seek to document the changes in adoption behavior that cause changes in energy savings.¹⁰⁴ It is important, therefore, to clearly articulate the assumed changes in the market, so they can be measured for the market effects study. If this is done properly the market effects evaluation can document changes in adoption, efficiency and provide an estimate of savings. This process also facilitates model specification.

A higher level of rigor is achieved when the market theory can be described in a narrative and/or a graphic logic model. A narrative or graphic logic model permits a greater depth of understanding of the indicators driving anticipated market outcomes. It can also help to identify the various sources of influence on market effects outside of the program efforts. The simplest approach to a logic diagram is to view the boxes as potential measurement indicators and the arrows as a hint to questions regarding causal links, program implementation theory, where to examine underlying behavioral change assumptions, and areas for researchable questions.

¹⁰¹ Sebold et al., pages 4-2 – 4-6.

¹⁰² TecMarket Works, pages 30-37.

¹⁰³ Nicholas P. Hall & John Reed. "Merging Program-Theory and Market-Theory in the Evaluation Planning Process." *Proceedings of the International Energy Program Evaluation Conference* (2001).

¹⁰⁴ Sebold et al., page 6-9, Figure 6-2.

Interviews or workshops should be used to develop the program theory. These should include program managers who understand the program purpose and can articulate the assumptions about how the program will change the market.¹⁰⁵

The key issues that should emerge from the workshops are program activities, identification of key market actors, assumed market barriers, expected outputs, outcomes and likely indicators of change, alternative hypotheses for change, external influences and causal links within anticipated timelines for achievement. Table 15 displays the Protocol for market theory and logic models.

An important distinction for the program theory/logic model development for a study under this Protocol is that the theory/logic model needs to be for a set of programs and capture both how the individual programs aim/hypothesize to affect the market as well as how they interact and support one another for market changes. The interaction, the degree to which and how to measure their ability to mutually support one another and how they interactively operate within a market are important analysis points and complications for a market level evaluation. Articulating the many assumptions this presents and then examining which are the most critical and how to test them are key to the degree to which the final market effects study will be comprehensive and defensible. A detailed understanding of how the market operates and how the various program interventions change or support that is a fundamental starting point for being able to attribute market changes to a group of programs, i.e., market effects.

Table 15. Required Protocol for Market Theory and Logic Models

Level of Rigor	Market Theory and Logic Model Requirements
Basic	Identification of assumptions about anticipated changes in the market and associated research questions. Market theory should include market operations and conditions, external influences, and assumptions about changes in the market (which could include market operational theory, market structure and function studies, and product and communication flows). Develop program theory and logic models across programs in that market. Analyze across both of these to examine program interventions, external influences and associated research questions. Theories and logic models should be generated through interviews with program staff and a sample of market actors.
Enhanced	Articulate market theory and, if reasonable, develop graphical model of market theory. Market theory should include market operations and conditions, and changes occurring in the market (could include market operational theory, market structure and function studies, and product and communication flows). Develop multiple program theory and logic models for those programs intervening in the market. Integrate the market theory and program theory/logic models to examine external and programmatic influences, assumptions about changes in the market and associated research questions. Theories and logic models should be generated through interviews or workshops with program staff from each of the programs and a sample of a wide variety of market actors. Use a literature review and other studies of these markets and iteration with program staff to ensure thoroughness in measuring the critical parameters for both market development from external influences and market effects.

¹⁰⁵ TecMarket Works, 30-38, 45-49, and 245-254. These sections of the *Evaluation Framework* (and the references provided for both chapters) will help in understanding the goals of program theory and logic models.

Determination of Indicators

The scoping study will determine what indicators should be used to assess market effects. The process emerges from the analysis of the market theory and program logic models but must also include an assessment of available data and primary data collection options. The use of smaller experimental designs imbedded into the program operations or quasi-experimental design is encouraged as a way to improve rigor without significantly increasing data collection costs over what is already required within this Protocol.

The market effects study should estimate what changes would have occurred in the market without program efforts. The indicators are used to draw conclusions about these changes. The focus should be on ultimate market indicators (the indicators of energy changes in the market). In developing the indicators there will be trade-offs that result in a level of rigor for the estimates. The scoping study should, therefore, address the level of rigor for the evaluation. The key considerations for the rigor of market effects estimates are the accuracy of the estimates of energy impacts and the accuracy of the attribution of market effects. The limitations of the market effects evaluation should be clearly articulated relative to these two issues and the scoping study should detail how the recommended approach addresses each.

The scoping study should also state the market assumptions and associated research questions to be addressed by the market effects study. At a market level, there are a variety of interventions that might occur and a variety of approaches that might be used to track and measure change in those interventions and their effect on the market. Table 16 indicates the general types of interventions that can influence change in a market and, therefore, suggests the types of indicators that might be tracked in market effects studies. The scoping study should clearly describe the relevance of the indicators to the market theory and how these indicators can be interpreted to indicate market effects.

Table 16. Types of Market Interventions and Associated Possible Indicators

Intervention Type	Ultimate Market Indicator	Other Indicators
Advertising/Outreach/Branding	Value of energy savings from sales and/or market share changes for targeted efficient measures	Awareness, source of awareness, intention to purchase, amount of exposure to intervention
Upstream Vendor Incentives	Value of energy savings from sales and/or market share changes for targeted efficient measures	Stocking practices, product availability, price
Trade Ally Training	Value of energy savings from sales and/or market share changes for targeted efficient measures or market share of efficient buildings	Surveys of practices, willingness to implement changes in installation or purchase, recommendation practices
End-user Training	Value of energy savings from sales and/or market share changes for targeted efficient measures or market share of efficient O&M practice	Surveys of practices, willingness to implement changes in operation or purchase
Downstream Incentives	Value of energy savings from sales and/or market share changes for targeted efficient measures	Non-participant awareness of the program, non-participant awareness of program participant experience

Detailed Market Effects Evaluation Plan

Once the scoping study is completed, the CPUC-ED will contract with an evaluator to implement the market effects evaluation. The first task for the evaluator will be to develop a detailed evaluation plan to implement the recommendations in the scoping study.

The *Evaluation Framework* discusses the need and value of an evaluation plan in some detail.¹⁰⁶ It is important that the Market Effects Evaluation Plan clearly documents the results of the scoping study and details the approach that should be taken to conduct the evaluation. The scoping study defines the market, details the market theory and logic and identifies the indicators to be used for tracking market effects. The evaluation plan captures these findings and details the process by which the indicator data will be collected or generated and describes the analysis approach to be used to estimate gross and net market effects and the resulting net energy market effects.

The evaluation plan should include a detailed description of the data collection approach including how indicators will be measured, population estimates and sampling targets. There should be a clear discussion of the analysis strategies and model specification if appropriate, and how the analysis plan will be developed. There should also be a schedule of milestones and deliverables and clear delineation of what information and data sources will be required from the utilities and other California entities.

¹⁰⁶ Ibid, 56-58.

Collection of Baseline and Longitudinal Indicators

Baseline studies are addressed in the *2001 Framework Study*¹⁰⁷ and the *Evaluation Framework*.¹⁰⁸ There are a variety of indicators that might be chosen to track market progress and thus determine whether market effects have occurred. Primary and secondary data are used for indicator studies. Primary data collection must be done carefully and samples used should be determined using the Sampling and Uncertainty Protocol. In those cases where secondary data exist, care should also be taken to understand the manner in which the data were collected to be certain of its appropriateness for market effects estimation. Where available, secondary data often provide a source for estimating market share for both efficient and non-efficient equipment sold in a market and can be the most effective way to obtain data for non-program affected areas.

Primary data collection involves collecting data (such as sales data) directly from actors in the market of interest. These types of studies vary in complexity, but at a minimum, the sample must be representative of the population of market actors. When surveying retailers and distributors, effort also needs to be made to adjust for double counting and to weight sales reports to account for total share of the market (see below). It is also possible to establish baselines for behaviors or energy using equipment by surveying end-users or market actors targeted for training or information services.¹⁰⁹ These types of studies all require that the questions asked enable the analyst to differentiate between sales of efficient and standard equipment or behaviors that improve efficiency over standard practice. Alternatives that provide potentially less biased or more readily accessible or controllable data should be examined. For example, changes in saturation over time might be a worthwhile alternative to sales data in some cases.

A higher level of rigor for primary data collection is achieved by carefully designed studies. For example, the California market share tracking studies for residential equipment are carefully designed to have a panel of participating retailers and distributors whose data can be weighted appropriately to estimate market share. To implement such a data collection effort requires establishing long-term relationships with retailers and distributors to provide sales data on a regular basis. Such studies require that the sample be carefully selected so that reported sales can be properly weighted to account for differential roles in the market by different retailers and distributors. Double counting also has to be avoided since distributors supply retailers. Highest levels of rigor are achieved by using multiple data sources to triangulate on the estimate of market share caused by the program efforts. Table 17 shows the Protocol for indicator studies.

¹⁰⁷ Sebold et al., pages 5-2 and 7-1 to 7-36.

¹⁰⁸ TecMarket Works, 254-262.

¹⁰⁹ Appliance sales have been tracked biennially in Wisconsin since 1993 (Energy Center of Wisconsin, 2004) by asking end-users about their purchases. The 2003 Appliance Sales Tracking Study is available at <http://www.ecw.org>. The Northwest Energy Efficiency Alliance tracks changes in behaviors for many of their programs by surveying representative samples of end-users and trade allies.

Table 17. Required Protocol for Market Effects Evaluation Indicator Studies

Level of Rigor	Indicator Study Requirements
Basic	Select appropriate market actor group for each indicator, survey representative samples of market actors able to report on each indicator from market experience. A baseline study must be conducted as early as possible. On-going tracking provides the basis for comparisons.
Enhanced	Select appropriate market actor group for each indicator. Conduct longitudinal study of representative samples of market actors able to report on each indicator from market experience. Samples weighted to represent known parameters in the population of interest. A baseline study must be conducted as early as possible, on-going tracking provides the basis for comparisons.

Analysis of Market Effects

The analysis of market effects has several components. First it should be determined if the indicators demonstrate any change in the market at all. This would be the estimation of gross market changes. Causality/attribution (which results in net market effects), sustainability and net energy impacts should then be estimated.

Gross Market Effects

Once the indicators have been collected for time one and time two, the analyst must determine the change in indicators across the time periods. For indicators such as market share and sales, it is reasonable to make direct comparisons. A variety of studies have shown that market share can be tracked directly overtime and these comparisons are fairly straightforward.

For other indicators such as awareness and knowledge, it is possible to make direct comparisons of indicators across time periods, but it is common that the direction and intensity of change in indicators will vary. One method that has been found to be effective in this type of situation is a binomial test.¹¹⁰

Estimating Causal Attribution

Causality should be examined to estimate net market effects. The goal of the activity is to estimate the proportion of market changes that can be attributed to program interventions using PGC and procurement funds, as versus those naturally occurring in the market or from interventions using non-PGC and non-procurement funds to arrive at market effects.

There are two primary approaches for estimating causal attribution, one uses a preponderance of evidence approach and the other uses a modeling approach. The ultimate goal for assessment of causal attribution is to avoid retrospective analysis in which contacts are asked to judge what efforts had effects on the market. Retrospective approaches have great potential for bias because contacts are themselves influenced and cannot maintain objective perspectives.

¹¹⁰ Richard F Spellman, Bruce Johnson, Lori Megdal and Shel Feldman, "Measuring Market Transformation Progress & the Binomial Test: Recent Experience at Boston Gas Company." *Proceedings, ACEEE 2000 Summer Study on Energy Efficiency in Buildings*. (Washington, DC: American Council for an Energy-Efficient Economy, 2000).

Preponderance of Evidence Approach for Attribution

In some cases, it is best to use a “preponderance of evidence” approach to assess the attribution of market effects. In this approach the analyst relies on triangulation from multiple data sources to draw conclusions about the presence and attribution of market effects. This approach is accomplished by interviewing and surveying knowledgeable market actors. Program staff, utility staff and trade allies provide useful information for understanding the context of sales and counts of behavior. Over time, these views provide much of the information needed to draw conclusions about attribution and sustainability. Systematic sampling is very important to ensure that bias is minimized.

A minimum level of rigor requires that samples of trade allies be included in the sampling plan, as they provide a less biased perspective due to their market-centric rather than energy efficiency-centric view. Rigor improves with more comprehensive samples of trade allies and other market actors. A variety of approaches can be used including choice and ranking surveys, focus groups, Delphi surveys and others.¹¹¹

The preponderance of evidence approach is inherently a qualitative analysis process in which the analyst uses multiple points of view to estimate the proportion of market effects that can be attributed to the program interventions. As noted above, the estimate will likely be a range, due in part to the qualitative nature of the analysis, but also to the difficulty in fully specifying all the factors that influence markets. The highest level of rigor relies on informants from multiple perspectives enabling the analyst to triangulate on the market effects. Table 18 shows the Protocol for the preponderance of evidence approach to attribution estimation.

Table 18. Required Protocol for Preponderance of Evidence Approach to Causal Attribution Estimation

Level of Rigor	Preponderance of Evidence Approach Requirements
Basic	A representative sample of market actors surveyed or interviewed to provide self-reports on perceived changes in the market, attribution and the sustainability of those changes.
Enhanced	Quasi-experimental or experimental design with comparison groups using a representative sample of market actors surveyed or interviewed to provide self-reports on perceived changes in the market, attribution and the sustainability of those changes.

Net Market Effects Modeling for Causation

The alternative to a preponderance of evidence approach is to use net effects modeling to control for non-PGC and non-procurement funded activities. In this approach the analyst uses multivariate models or simultaneous modeling systems to estimate net market effects. A variety of methods can be used. Some of these are discussed in Chapter 7 of the *2001 Framework Study*, in which it is suggested that the use of dynamic baselines in which a forecast of market changes are made in time one, using time one data, and then in time two the forecast is tested against the

¹¹¹ Sebold et al., 6-23 - 6-25 and 7-5 - 7-7.

actual conditions of time two.¹¹² Additional methods are being explored and hold a great deal of promise for clarifying the extent of market effects caused by energy efficiency program efforts.

A modeling approach permits the analyst to specify a model of the program theory and to test that model with data gathered in time one and time two. This is a growing area of investigation with a limited number of studies having been completed as of 2005. In constructing such a model, it is important that the model specifications reflect the complexity of the market. This is the greatest challenge for this approach. It is likely that such an approach will require multiple equations to model the various activities that occur in a market and the various points of intervention that energy efficiency programs exert on a market.

Given the early stage of development for this type of approach, it is not possible to determine levels of rigor. Advancements on these methods are being developed as it appears this approach could offer a greater level of rigor, quantification and testing than prior methods.

Estimating Sustainability

Sustainability is the degree to which one can expect the market changes to last into the future. Sustainability is not readily estimated using net effects modeling therefore the preponderance of evidence approach is the most frequently used for estimating sustainability. As with attribution a minimum level of rigor requires that samples of trade allies be included, as they provide a less biased perspective due to their market-centric rather than energy efficiency-centric view. Rigor improves with more comprehensive samples of trade allies and other market actors. A variety of approaches can be used including choice and ranking surveys, focus groups and Delphi surveys.¹¹³ Another valid approach to estimate sustainability is identifying changes in market structure and operations, and how the changed market contains mechanisms to sustain them. This could include examining profitability analyses for important support businesses or business operations and how these are maintained without continued program intervention.

As noted previously, the preponderance of evidence approach is inherently a qualitative analysis process in which the analyst uses multiple points of view to estimate whether the market effects attributed to the program interventions can be expected to continue into the future. The highest level of rigor relies on informants and analyses from multiple perspectives enabling the analyst to triangulate on sustainability. A market with multiple support areas for continued sustainability will have a greater likelihood of having the changed market operation be sustainable.

The result of the estimation of sustainability is a statement on the likelihood of the market effects continuing without the energy efficiency program intervention or with reduced interventions. Given California's current interest in market effects that have recently occurred rather than those forecasted to occur (the focus of market effects estimation efforts in several other states), there is significantly less need for measures of sustainability. This issue, however, is a critical one whenever forecasts of market effects are the dominant evaluation concern.

Market Effects Metrics and Energy Savings

When the net effects modeling or preponderance of evidence approach is used to estimate net market effects, the analysis will result in an estimate of market share for the sales or counts of

¹¹² Sebold et al., pages 7-1 to 7-36.

¹¹³ Ibid, 6-23 to 6-25 and 7-5 to 7-7.

behavior, or other indicator(s) attributed to the program. The net market effects must then be linked to an estimate of energy savings. The sales and counts of behavior or other indicators used to estimate market effects are linked to the energy savings for those measures or behaviors estimated through the M&V or Impact Evaluation Protocols, or in DEER. Savings estimates are directly applied to net changes in sales, counts of behavior and the like by multiplying the savings term by the associated amount of energy usage for the equipment or behavior of interest.

In some cases when net market effects modeling is used, it is anticipated that energy will be a term in the equation. Therefore, rather than linking the estimated net market effect to a savings estimate, the analyst will use the energy term as the dependent variable that is being modeled. The indicators will be the independent variables specified to explain the energy term.

Reporting

The evaluation report should also address the level of rigor for the study. The key considerations for the rigor of market effects estimates are the accuracy of the energy impact estimates and the accuracy of the attribution of market effects. The limitations of the market effects evaluation should be clearly articulated relative to these two issues.

In addition to estimating net energy market effects, the market effects evaluation report should clearly state the market assumptions and associated research questions addressed by the market effects study. The market effects evaluation report should clearly articulate the logic of the approach - whether using a preponderance of evidence approach to justify net market effects or a regression-based modeling approach.¹¹⁴ Both approaches should build on the market theory as a hypothesis that was developed earlier in the scoping study.

Market effects evaluations will result in a report documenting the evaluation and its findings. The Reporting Protocol describes the content of the market effects evaluation report. The key aspects of that report include the following:

- Documentation of the market theory and the program theory/logic model as developed in the scoping study including an assessment of the initial market theory and program theory based on the results of the evaluation, and recommendations for a revised market theory/program theory, if needed;
- Documentation of the data collection and analysis process used for the market change indicators, whether the data used were primary data or secondary data. What indicators were used, how the data were assembled, collected and analyzed and the results of the various indicators studies;
- Documentation of the estimation of gross market effects that result from an analysis of the indicators, regression modeling or triangulation of the two;
- Documentation of the process used and results obtained for estimating causal attribution and sustainability and the resulting estimate of net market effects; and

¹¹⁴ Modeling the market processes and change processes, some sequential and some simultaneous, is encouraged as an enhancement for a regression-based modeling approach over a single-equation model. Any use of a single-equation model must justify the model specification and its ability to capture the critical evaluation elements seen in the market theory and program theory/logic models.

- Documentation of the process used and results obtained when estimating net energy market effects. What energy data were used and how they were linked to the estimate of net market effects.

Guidance on Skills Required to Conduct Market Effects Evaluations

This Protocol suggests that there are two primary strategies for conducting net market energy effects evaluations, each of which can be considered rigorous when well-executed.

A preponderance of evidence approach, in which the analyst relies on triangulation from multiple data sources, is used to draw conclusions about the presence of market effects. While secondary data can be used in this approach, significant primary data collection is expected. The preponderance of evidence approach, therefore, requires skills in designing and implementing survey and interview instruments to collect indicators that correspond to the theory and reflect how the market is thought to operate. The evaluators should have sufficient experience to implement surveys, interviews, group interviews and other types of primary data collection activities. Since energy savings are drawn from impact evaluation results, the firms conducting market effects evaluations should have vast experience in energy efficiency markets, the social sciences, and interview and survey instrument design, implementation and analysis.

The net market effects modeling approach, in which the analyst uses multivariate models to estimate net market effects, can use primary or secondary data, although the use of secondary data has been most common. This type of approach is largely dependent on professional evaluators experienced in regression-based and multivariate modeling. The evaluator must be able to specify a model of the market during the scoping study and then populate the model with secondary and primary data. One approach is to develop a forecast of the market in time one and then test it in time two. Another approach is to take a retrospective approach using secondary data over a multi-year period. Other approaches are still emerging. The major limitation of the net market effect modeling approach is the availability of sufficient data to meet the model requirements. Modeling systems and/or specification that can mirror market operations and program theory interventions have still to be developed. Modeling the market processes and change processes, some of which are sequential while some are simultaneous, is an enhancement to a regression-based modeling approach and is encouraged over a single-equation model. Any use of a single-equation model must justify the model specification and its ability to capture the critical evaluation elements seen in the market theory and program theory/logic models. Thus, a scoping study should be used to determine if such an approach is warranted and can be expected to be successful.

Considerations for Conducting a Market Effects Evaluation

The key consideration for conducting a market effects evaluation is determining whether market level effects are expected. As noted previously, programs that operate within a market have ripple effects on other programs operating in that market. Obvious examples are how the United States Department of Energy and the Environmental Protection Agency's ENERGY STAR efforts interact with the California energy efficiency program activities to encourage the adoption of energy-efficient appliances in the residential sector. All states are showing increased adoption of ENERGY STAR appliances, but the question remains as to what part of this market change is

induced by the California programs. Another example concerns new residential and commercial construction activities that are affected by the implementation of California's building codes (Title 24) and by California's program activities designed to change construction behaviors and code-covered practices.

A market effects evaluation is appropriate when net market effects are used to justify a program or group of programs,¹¹⁵ or when net market effects may be of interest to the Joint Staff for a set of programs operating in the same market. At the same time, the Impact Evaluation Protocol does not measure non-participant spillover due to the assessment that these are best measured by market effects evaluations. This means that the full effect of California's investments in energy efficiency programs may not be obtained through the sum of individual program evaluations but, instead, through an analytic derivation from the program-level evaluations and the market effects evaluations. (See the discussion in the Impact Evaluation Protocol on interaction with the Market Effects Protocol.) Market effects evaluations are needed, then, to have the information to derive the full impacts of the California efforts and investment.

As noted previously, determining the "market" is an important step in the scoping study. The Joint Staff will make recommendations for markets in which they expect market effects to be measured. Markets of possible interest to the Joint Staff include, but are not limited to:

- Residential appliance;
- Mass marketing campaigns;
- Residential construction;
- Nonresidential construction;
- Agricultural services;
- Commercial lighting;
- Residential lighting;
- Education (general public and targeted groups, e.g., contractors);
- Training programs; and
- Technical assistance programs.

Summary of Protocol-Driven Market Effects Evaluation Activities

1	Joint staff identifies the markets or market sectors (and the associated set of programs) that will receive a market effects evaluation and identifies the potential approach and rigor level for the scoping study.
2	Joint staff identifies market- or market sector-specific study needs that will be assessed (including program-specific or program group specific study needs) from the evaluation. CPUC-ED issues request for proposals for market effects scoping study, selects the scoping study contractor and establishes a scope(s) of work.

¹¹⁵ TecMarket Works, 247, Figure 10.1.

3	Evaluation contractor develops scoping study. A scoping study will more finely define the market boundaries for the study, including its location, the utilities involved, the equipment or behaviors to be assessed and the program-influenced years of interest. The scoping study will develop a market theory and a logic model; identify the market change indicators to track; and the available primary and secondary data sources. The study will also identify the hypotheses to test and the data collection approach, and provide a recommended analysis approach and model specification (if appropriate).
4	A market change theory and logic model (MCT/LM) should be developed to identify assumed direction of effects and indicators for measuring effects. The market theory should include market operations and conditions, and changes occurring in the market (could include a market operations theory, market structure and function scenarios, and product and communication flows) The theory and logic model should be generated through interviews or workshops with program staff from each of the programs that are expected to influence the market being assessed and a sample of a wide variety of market actors and should incorporate a literature review.
5	Joint staff reviews the scoping study and determines how to proceed with the Market Effects Evaluation. CPUC-ED issues request for proposals for evaluation contractors, selects the contractor, establishes a final scope(s) of work and negotiates the contract.
6	All market effects evaluation teams must be staffed to meet the skills required for the research design, sampling, appropriate and selected evaluation method, uncertainty analysis and reporting requirements.
7	A research design and sampling plan should be developed to meet Protocol requirements at the market level to meet the Joint Staff assigned study rigor level. This includes meeting requirements from the Sampling and Uncertainty Protocol and the Reporting Protocol, as applicable. The evaluation contractor will develop an Evaluation Plan, submit it to the CPUC-ED and revise as necessary.
8	Indicators studies conducted as part of the Market Effects Evaluation should be based on the results of the scoping study, address the appropriate market actor group(s) for each indicator.
9	All Market Effects Evaluations must meet the requirements of the Sampling and Uncertainty Protocol. The 90/10 level of precision is a minimum precision target for the most important data collection efforts on its most important variables. Which data collection efforts and variables are considered to be the most important will be determined in close collaboration with the CPUC-ED
10	The gross market effects and the estimate of energy savings associated with the market effects should be estimated. Estimation of gross market effects can be as simple as comparing indicators between time one and time two and then multiplying the energy value derived in an M&V supported impact assessment or from DEER, or using a CPUC-ED-approved net energy effects model.
11	Attribution or causality should be addressed to estimate net effects using either a preponderance of evidence approach or a net effects modeling approach. a. For a preponderance of evidence approach a determination of attribution should use quasi-experimental or experimental design with comparison groups using a representative sample of market actors. This may include interviews to provide self-reports on perceived changes in the market, attribution and the sustainability of those changes as well as direct observation or other data to support changes resulting from the program.

	b. For a net effects modeling approach to estimate causality, the model specifications must reflect the complexity of the market. It is likely that such an approach will require multiple equations to model the various activities that occur in a market and the various points of intervention that energy efficiency programs exert on a market.
12	Sustainability should be addressed using a preponderance of evidence approach.
13	Develop draft evaluation report to include meeting all requirements in the Reporting Protocol and incorporating the program's performance metrics.
14	Develop final evaluation report in accordance to guidance provided by Joint Staff.
15	Submit final evaluation report to the CPUC-ED.
16	Once the report is accepted by the CPUC-ED, develop abstracts and post them and the report on CALMAC Web site following the CALMAC posting instructions

Note: The steps included in this evaluation summary table must comply with all the requirements of the Market Effects Evaluation Protocol.

Sampling and Uncertainty Protocol

Introduction

There are some important similarities between the pre-1998 protocols and the 2006 Protocols related to impact and M&V studies. Both sets of protocols focus on obtaining reliable estimates of energy and demand impacts. Reliable estimates are interpreted as estimates that are reasonably accurate and precise, that is, they contain a minimal amount of error from a variety of sources such as sampling error, measurement error, and model misspecification error. The pre-1998 protocols concern the same issues listed in the *Evaluation Framework*:¹¹⁶

- Non-response and other forms of selection bias;
- Measurement error;
- Erroneous specification of the statistical model;
- Choosing an inappropriate baseline;
- Self-selection of program participants;
- Misinterpretation of association as causal effects;
- Construct validity;
- Statistical validity;
- Internal validity; and
- External validity.

However, the two protocols also have differences, the two primary of which relate to the number of study types and the degree of precision required for energy-use estimates. The 2006 Protocols must address an additional set of studies that include process evaluations, indirect impact evaluations for education, training and advertising programs, and market effects evaluations. The reliability of information produced by these studies is equally important and must be addressed in the 2006 Protocols.

The pre-1998 protocols require 90/10 precision for estimates of annual energy use while the 2006 Protocols set precision targets¹¹⁷ whenever possible for a variety of parameters including savings.¹¹⁸ Precision targets are set rather than required since, as discussed in the *Evaluation Framework* and its cited study of this issue by Sonnenblick and Eto, bias could be much more

¹¹⁶ See *Evaluation Framework*, 292-294 for examples and definitions of the terms listed here, along with citations to reference documents.

¹¹⁷ A precision target is a goal established at the beginning of an evaluation based in large part on initial estimates of uncertainty. If an evaluator fails to actually achieve the targeted level of precision, there will be no penalties since the assumptions underlying the sample sizes proposed in each evaluation plan will have been *clearly presented and carefully documented*. A failure to meet the precision target for a given program will only require an adjustment of the input assumptions prior to the next evaluation cycle and, if necessary, a reallocation of evaluation dollars to support increased sample sizes.

¹¹⁸ The *Evaluation Framework* proposed no precision targets or requirements for savings or for any other parameters associated with such studies as process and market effects evaluations.

important than precision for the reliability of the savings estimates or the cost-effectiveness calculations.¹¹⁹ In addition, as any evaluation study proceeds, the data collected could contain much more error than originally thought, requiring more resources to be devoted to reducing this bias and fewer resources devoted to achieving the required statistical precision. Or, the variability in the savings could be so great that it would be impossible to meet the precision requirement. The evaluator must have the flexibility to respond to data issues as they arise in order to maximize the reliability of the savings. Therefore, focusing on sample error, while giving relatively little attention to these other sources of error, would compromise the CPUC's objective of obtaining *reliable* estimates of kWh and kW impacts.

Finally, the guidelines regarding sampling and uncertainty must be followed for each utility service territory. For example, precision targets, when specified for a particular level of rigor, must be set for *each* utility service territory.

Precision: Gross and Net Impact, Measurement and Verification, and Verification Activities

There are a number of impact-related activities concerning precision addressed in this section:

- Estimation of gross impacts (including M&V);
- Estimation of net impacts;
- M&V in support of specific measure studies; and
- Verification studies in support of non-Impact Evaluation Protocol gross and net impacts.

The issue of precision for each of these types of analytical studies is addressed in Table 19 through Table 23.

¹¹⁹ California Framework, p. 296.

Table 19. Required Protocols for Gross Impacts¹²⁰

Rigor Level	Gross Impact Options
Basic	Simplified Engineering Models: The relative precision is 90/30 ¹²¹ . The sampling unit is the premise. The sample size selected must be justified in the evaluation plan and approved as part of the evaluation planning process.
	Normalized Annual Consumption (NAC) Models: There are no targets for relative precision. This is due to the fact that NAC models are typically estimated for all participants with an adequate amount of pre- and post-billing data. Thus, there is no sampling error. However, if sampling is conducted, either a power analysis ¹²² or justification based upon prior evaluations of similar programs must be used to determine sample sizes. The sample size selected must be justified in the evaluation plan and approved as part of the evaluation planning process.
Enhanced	Regression: There are no relative precision targets for regression models that estimate gross energy or demand impacts. Evaluators are expected to conduct, at a minimum, a statistical power analysis as a way of initially estimating the required sample size. ¹²³ Other information can be taken into account such as professional judgment and prior evaluations of similar programs. The sample size selected must be justified in the evaluation plan and approved as part of the evaluation planning process.
	Engineering Models: The target relative precision for gross energy and demand impacts is 90/10. The sampling unit is the premise. The sample size selected must be justified in the evaluation plan and approved as part of the evaluation planning process.

¹²⁰ See the Impact Evaluation Protocol for a description of methods and page references in the *Evaluation Framework* for further information and examples.

¹²¹ Also of interest, in addition to the relative precision, are the actual kWh, kW, and therm bounds of the interval.

¹²² Statistical power is the probability that statistical significance will be attained, given that there really is a treatment effect. Power analysis is a statistical technique that can be used (among other things) to determine sample size requirements to ensure statistical significance can be found. Power analysis is only being required in the Protocol for determining required sample sizes. There are several software packages and calculation Web sites that conduct the power analysis calculation. One of many possible references includes: Cohen, Jacob (1989) *Statistical Power Analysis for the Behavioral Sciences*, Lawrence Erlbaum Associates, Inc.

¹²³ Ibid.

Table 20. Required Protocols for Net Impacts

Rigor Level	Net Impacts Options
Basic	For the self-report approach (Option Basic.1), given the greater issues with construct validity and variety of layered measurements involved in estimating participant NTGRs, no relative precision target has been established. ¹²⁴ To ensure consistency and comparability a minimum sample size of 300 sites (or decision-makers in cases where decision-makers cover multiple sites) or a census ¹²⁵ , whichever is smaller, is required.
Standard	If the method used for estimating net energy and demand impacts is regression-based, there are no relative precision targets. If the method used for estimating NTGRs is regression-based (discrete choice), there are no relative precision targets. In either case, evaluators are expected to conduct, at a minimum, a statistical power analysis as a way of initially estimating the required sample size.¹²⁶ Other information can be taken into account such as professional judgment and prior evaluations of similar programs. For the self-report approach (Option Standard.2), there are no precision targets since the estimated NTGR will typically be estimated using information collected from multiple decision-makers involving a mix of quantitative and qualitative information around which a standard error cannot be constructed. Thus to ensure consistency and comparability, for such studies, a minimum sample size of 300 sites (or decision-makers in cases where decision-makers cover multiple sites) or a census, whichever is smaller, is required.
Enhanced	The requirements described for Enhanced apply depending on the methods chosen.

¹²⁴ This is considered the best feasible approach at the time of the creation of this Protocol. Like the other approaches to estimating the net-to-gross ratio (NTGR), there is no precision target when using the self-report method. However, unlike the estimation of the required sample sizes when using the regression and discrete choice approaches, the self-report approach poses a unique set of challenges to estimating required sample sizes. These challenges stem from the fact that the self-report methods for estimating free-ridership involve greater issues with construct validity, and often include a variety of layered measurements involving the collection of both qualitative and quantitative data from various actors involved in the decision to install the efficient equipment. Such a situation makes it difficult to arrive at a prior estimate of the expected variance needed to estimate the sample size.

Alternative proposals and the support and justifications that address all of the issues discussed here on the aggregation of variance for the proposed self-report method may be submitted to Joint Staff as an additional option (but not instead of the Protocol requirements) in impact evaluation RFPs and in Evaluation Plans. Joint Staff may elect to approve an Evaluation Plan with a well-justified alternative.

¹²⁵ A census is rarely achieved. Rather, one *attempts* to conduct a census, recognizing that there will nearly always be some sites, participants or non-participants who drop out for a variety of reasons such as refusals or insufficient data.

¹²⁶ Statistical power is the probability that statistical significance will be attained, given that there really is a treatment effect. Power analysis is a statistical technique that can be used (among other things) to determine sample size requirements to ensure statistical significance can be found. Power analysis is only being required in the Protocol for determining required sample sizes. There are several software packages and calculation Web sites that conduct the power analysis calculation.

Table 21. Required Protocols for Measure-level Measurement and Verification

Rigor Level	M&V Options
Basic	Simplified Engineering Models: The target relative precision for gross energy and demand impacts is 90/30. The sample unit may be the individual measure, a particular circuit or point of control as designated by the M&V plan.
Enhanced	Direct Measurement and Energy Simulation Models: The target relative precision for gross energy and demand impacts is 90/10. The sample unit may be the individual measure, a particular circuit or point of control as designated by the M&V plan.

Table 22. Required Protocols for Sampling of Measures Within a Site

The target relative precision is 90/20 for each measure selected for investigation. The sampling unit (measure, circuit, control point) shall be designated by the M&V plan. The initial assumption regarding the coefficient of variation for determining sample size is 0.5.

Table 23. Required Protocols for Verification

Rigor Level	Verification Options
Basic	The target relative precision is 90/10. The key parameter upon which the variability for the sample size calculation is based is binary (i.e., Is it meeting the basic verification criteria specified in the M&V Protocol?).
Enhanced	The target relative precision is 90/10. The key parameter upon which the variability for the sample size calculation is based is binary (i.e., Is it meeting the enhanced verification criteria specified in the M&V Protocol?).

Of course, when sampling from any population it should always be assumed that there will be some attrition due to such factors as refusals to participate in a telephone survey or an on-site inspection, or insufficient data. As a result, a larger sample than is actually needed should always be drawn based on the best estimate of expected attrition.

Development of the Evaluation Study Work Plan

For each study in the evaluator's defined set of studies, the evaluator must prepare a detailed evaluation work plan (plan) that allocates resources to maximize reliability for the program group and takes into account that the level of rigor will likely vary by program. In many cases, the evaluator will be required to develop a separate work plan for each program in the study set. In some cases, a draft plan will be required as part of the initial proposal package, in others the evaluator may be required to develop this work plan after the hiring process is complete. As part of this plan, the evaluator must specifically address the various sources of error that are relevant and explain how the resources allocated to each will mitigate the error¹²⁷. They must also estimate the statistical precision that the planned evaluation will achieve. It is also recognized

¹²⁷ In the pre-1998 M&E Protocols, there was no requirement to address these sources of error in the research plan. Evaluators only had to describe in the final report whether they had to address these various errors and, if so, what they did to mitigate their effects.

that the targeted precision at the *program level* must be allowed to vary in ways that produce the greatest precision at the *program group level*. For example, in some cases accepting a lower level of precision for programs with small savings might allow for the allocation of greater resources to programs with larger savings, thus increasing the achieved precision for the program group.¹²⁸

The Joint Staff and other outside resources as deemed appropriate by the CPUC will review the evaluation plan submitted and discuss with the independent evaluator any changes they deem necessary to maximize the reliability of the savings estimates at the program group level.¹²⁹ The Joint Staff might decide to increase the sample size in order to increase precision, recognizing that other sources of error will receive fewer resources, or they might decide to reduce the sample and settle for lower precision in exchange for a greater effort to reduce non-response bias. In the final plan, the evaluators and Joint Staff will endeavor to allocate their available evaluation resources in such a way as to maximize the reliability of the savings and the value received from the evaluation efforts. In order to more adequately address accuracy and/or precision, once evaluation studies are underway, Joint Staff may adjust the allocation of resources that were initially dedicated to the evaluation of a given program, program group, or study set.

The level of rigor assigned to each program will vary depending on the evaluation priorities and budgets discussed above. However, because each program is somewhat unique with respect to the various sources of bias, there is no specific set of required methods and level of effort for minimizing bias that can be assigned based on the level of evaluation rigor assignment.

At the same time, every impact and indirect impact evaluation plan, analysis and report is expected to seriously address, at a minimum, each and every one of the ten sources of uncertainty listed in the introduction of this section. The assessment of the potential issues, testing, minimization approaches and mitigation efforts are to be discussed in the evaluation plan and carried forward through the evaluation and evaluation reporting. This assessment and reporting needs to include the justification based on prior evaluations, evaluation science and other research (with appropriate citation) that support the evaluation research design decisions made in the evaluation plan and the handling of the issues through the analyses. The reporting should include specific data collection, measurement and handling of each issue at a level of detail that allows the study results to be replicated. Results from tests of alternative methods of data handling should be included. For example, if outliers are dropped from the analysis, the reporting should include the methods used to identify outliers, analysis results with and without outliers, and the justification used in deciding to remove some or all of the outliers. Data cleaning methods and decision rules should be supplied with at least some testing of the analysis impacts produced by varying the primary parameters in these decision rules. Similarly, any sampling and site selection parameters need to be examined for potential bias with appropriate research questions and tests being conducted on key parameters.

¹²⁸ See *California Framework*, pp. 305-313 for a description and some examples of how to allocate resources and sample sizes to obtain the smallest possible error bound for a group of programs.

¹²⁹ *Ibid*, 298-300 for a description of calculating error bounds and precision levels for different types of evaluation study integrations.

Process Evaluations

For process evaluations, the focus is on reliability at the program level, with the level of evaluation rigor varying as a function of evaluation priorities and budgets. However, because each program is somewhat unique, with respect to the data being collected and the various sources of bias, there is no specific set of required methods and level of effort for minimizing bias that can be assigned to a program that has been assigned a given level of evaluation rigor.

Requiring 90/10 precision, for example, for all inquiries is very likely infeasible and not cost-efficient because budgets are limited, there is often a large set of evaluation questions to be addressed (i.e., many different questions and parameters for which some level of precision could be desired), not all of which are quantitative, and the information sought from different survey and interview groups might not be equally valuable. For example, one might want to field a small survey to get a sense of the motivation of a particular market actor. Again, it is important for the evaluator to have the flexibility to maximize the reliability of their findings. However, the 90/10 level of precision should be adopted as a minimum precision target for the most important data collection efforts on its most important variables. Which data collection efforts and variables are considered to be the most important for process evaluations will be determined by the independent evaluator in close collaboration with utility EM&V staff.

There are circumstances when it might be desirable to use M&V as input to the analysis of a problem being investigated in a process evaluation. If M&V is not conducted by the Joint Staff evaluations, utility evaluation staff may choose to specify M&V activities within the process evaluation RFP.¹³⁰ If the M&V Protocol is used for purposes outside impact, indirect impact and verification analysis, a target precision should, at a minimum, be 30 percent precision at a 90 percent confidence level (or 90/30 precision).

The evaluator must prepare a detailed plan that allocates resources in order to maximize reliability for the findings and for key parameter estimates for each program in the group. As part of this plan, the evaluator must specifically address the various sources of error that are relevant and explain how the resources allocated to each will minimize and/or mitigate the error.¹³¹ They must also estimate the statistical precision that the planned evaluation will achieve on selected primary quantitative measurements.

The Joint Staff and other outside resources as deemed appropriate by the CPUC will review the evaluation plan submitted and discuss with the independent evaluator any changes they deem necessary to maximize the reliability of the findings at the program level. The evaluation staff might decide to increase the sample size in order to increase precision, recognizing that the other sources of error will receive fewer resources. Or it might decide to reduce the sample size and settle for lower precision in exchange for a greater effort to reduce non-response bias. In the final plan, evaluation resources will be allocated in a way that maximizes the reliability of the findings for each program.

¹³⁰ Coordination of M&V studies for process and impact purposes is a key issue that must be addressed by the evaluation plans for both process and impact evaluation.

¹³¹ In the pre-1998 Protocols, there was no requirement to address these sources of error in the research plan. Evaluators only had to describe in the final report whether they had to address these various errors and, if so, what they did to mitigate their effects.

Market Effects

The focus is on the market level for market effects evaluations. The level of rigor assigned to a particular market effects study will depend on the evaluation priorities and budgets. However, because each market effects study will be somewhat unique with respect to the data being collected and the various sources of bias, there is no specific set of required methods and level of effort for minimizing bias that can be assigned to a given market effects study.

Requiring 90/10 precision for all estimates, for example, is very likely infeasible and not cost-efficient because budgets are limited, there are often a large set of evaluation questions, outcomes and causal mechanisms to be assessed in a market effects evaluation (i.e., many different questions and parameters for which some level of precision could be desired), and the information sought from different survey, interview groups and data sources might not be equally valuable. For example, one might want to field a small survey to roughly estimate the number of HVAC contractors who actively promote energy-efficient air conditioners. Again, it is important for the evaluator to have the flexibility to maximize the reliability of their findings. However, the 90/10 level of precision should be adopted as a minimum precision target for the most important data collection efforts on its most important variables. Which data collection efforts and variables are considered to be the most important will be determined by the independent evaluator in close collaboration with the CPUC.

The evaluator must prepare a detailed evaluation plan that allocates resources in order to maximize reliability of market-level estimates. As part of this plan, the evaluator must specifically address the various sources of error that are relevant and explain how the resources allocated to each will minimize and/or mitigate the error (e.g., non-response bias, measurement error, and self-selection bias).¹³² They must also estimate the statistical precision that the planned evaluation will achieve on key estimates and for the overall estimate of market effects (to include the propagation of error).

The Joint Staff and other outside resources as deemed appropriate by the CPUC will review the evaluation plan submitted and discuss with the independent evaluator any changes they deem necessary to maximize the reliability of the estimates at the market level. For example, The Joint Staff might decide to increase the sample size or budget in order to increase precision for specific parameters or study elements, recognizing that the other sources of error will receive fewer resources. Or it might decide to reduce the sample and settle for lower precision in exchange for a greater effort to reduce non-response bias. In the final plan, evaluation resources will be allocated in a way that maximizes the reliability of the market-level estimates.

System Learning

The hallmark of any learning system is that feedback is processed and any necessary course corrections are made. Once a particular evaluation is launched, it's certainly possible that mid-course adjustments will be made to the initial plan to maximize savings reliability. For example,

¹³² In the pre-1998 Protocols, there was no requirement to address these sources of error in the research plan. Evaluators only had to describe in the final report whether they had to address these various errors and, if so, what they did to mitigate their effects.

the coefficients of variation (CVs)¹³³ for certain key parameters, measures, end-uses or programs might actually be smaller than anticipated or the random and/or systematic measurement error might be worse. As data are collected and assessed, decisions can be made regarding the reallocation of resources.

Once a particular study is completed or all the studies within a given group are completed, the CPUC-ED, utility EM&V staff and the independent evaluators can review the achieved precision and the results of efforts to minimize bias and recommend how evaluation resources can be reallocated for the next evaluation cycle.

Acceptable Sampling Methods

It is rarely possible, for a variety of different reasons, to conduct a census of any population (e.g., program participants, programs non-participants or lighting vendors).¹³⁴ Especially in a state the size of California, this is due largely to the fact that many of the populations are quite large and the cost of attempting a census study would be prohibitive. Instead, random samples drawn from these populations are almost always used as a way to estimate various characteristics of these populations. The specific approaches to maximizing precision are left up to the independent evaluator. For example, one can choose from a variety of sample procedures recognized in the statistical literature, such as sequential sampling, cluster sampling, multi-stage sampling and stratified sampling with regression estimation. There are many available books on sampling techniques that can be used as reference.¹³⁵

Skills Required for Sampling & Uncertainty

Population database work and simple random sampling (or census) do not require an advanced statistics background. Other more complex sample designs require basic training and/or experience in statistics to ensure that the methods are understood and applied correctly. Those conducting and reviewing this work should have at least basic graduate statistics or equivalent experience with a mentor in this area. The skills required for addressing the uncertainty associated with the various methods for estimating the gross and net energy and demand impacts as well as the net impacts are described as part of the Impact Protocols.

Audience and Responsible Actors

- **Program Evaluators** should understand and implement this Protocol. They also need to be able to correctly estimate the expected precision and accuracy. Based on the achieved precision and accuracy, they must recommend any reallocation of evaluation resources going forward;
- **CPUC-ED CEC and Utility EM&V Staff** should understand this Protocol and be able to correctly interpret the expected and achieved levels of precision and accuracy in order to

¹³³ The sample standard deviation divided by the sample mean. See page 320 of the *Evaluation Framework*.

¹³⁴ In process evaluations, a census is possible in some more limited populations such as staff and program contractors.

¹³⁵ The two cited in the *Evaluation Framework* are 1) Cochran, William G. *Sampling Techniques*. New York: John Wiley & Sons, 1977 and 2) Sarndal, Carl-Eric, Bengt Swensson and Jan Wretman. *Model Assisted Survey Sampling*. New York: Springer-Verlag, 1992.

accept or reject any recommendations regarding the reallocation of evaluation resources going forward; and

- **Utility System Planners** should be able to understand the achieved precision and accuracy and the overall reliability of the savings in order to assess their resource value.

Key Metrics and Information Covered

All evaluation reports must contain a variety of information regarding the sample design and implementation as well as a variety of information regarding the various sources of bias encountered and efforts to mitigate them. These are outlined below.

Sample Size and Precision

Whenever estimates are based on a sample in any evaluation, the following information, as appropriate, must be reported:

- The definition of the population from which the sample was drawn;
- The sample design (e.g., simple random, stratified random and two-stage);
- The assumptions and related documentation upon which the initial sample size calculations were based (e.g., CV for key inputs in an engineering algorithm, CV for proportion of audit participants who adopt recommendations, the specified statistical power, effect size, confidence level and alpha level);
- The details of how the initial sample sizes were calculated to achieve the agreed upon level(s) of precision;
- The achieved precision around program-level gross and net kWh, kW, and therm impacts, key process evaluation measurements, and other program impacts such as attitude change and knowledge gains;
- The confidence intervals specified in terms of the kWh, kW, and therm impacts;
- The details of how the achieved sample size was used to calculate the precision; and
- Response rate and attrition and any suspected non-response bias and efforts to address it.¹³⁶

Validity and Research Design

- Discuss threats to internal validity (the extent to which alternative potential causes of the measured effect have been ruled out within the analysis);¹³⁷
- Discuss threats to external validity (the extent to which the analysis results found for a sample are true for the population and the program overall);¹³⁸ and

¹³⁶ See "Definitions of Response, Refusal, and Cooperation Rates" prepared by the Council for Marketing and Opinion Research and the "CASRO Guidelines for Survey Research Quality," prepared by the Council of American Survey Research Organizations (www.casro.org).

¹³⁷ See *Evaluation Framework*, pp. 292-295 and 425.

¹³⁸ Ibid, 292-295 and 421.

- Discuss assessment of construct validity and potential remaining issues of construct validity for the primary evaluation outputs (the extent to which the measurement (and instrumentation, such as survey wording) captures the underlying abstract idea).¹³⁹

Accuracy

Regression models:

- Describe procedures used for the treatment of outliers, missing data points and weather adjustment;
- Describe what was done to control for selection bias, if suspected;
- Describe what was done to control for the effects of background variables, such as economic and political activity that may account for any increase or decrease in consumption in addition to the program itself;
- Describe procedures used to screen data for inclusion into the final analysis dataset. Show how many customers, installations or observations were eliminated with each screen. The reviewer should be able to clearly follow the development of the final analysis dataset;
- Regression statistics: For all final models, provide standard regression statistics in a tabular form;
- Specification: Refer to the section(s) of the study that present the initial and final model specifications that were used, the rationale for each, and the documentation for the major alternative models used. In addition, the presentation of the specification should address, at a minimum, the following issues:
 - Describe how the model specification and estimation procedures recognize and address heterogeneity of customers (i.e., cross-sectional variation);
 - Describe how the model specification and estimation procedures recognize and address changes in factors that affect consumption over time (i.e., time series variation), apart from program effects;
 - Describe how the model specification and estimation procedures recognize and address the fact that participants self-select into that status, and discuss the effects of self-selection on model estimates whether or not self-selection is treated explicitly;
 - Describe how truncation within the data and regression towards the mean within the participant population (e.g., within low-income populations) is tested for, the results of this test, and how model specification and estimation procedures recognize and address these issues;
 - Discuss the factors, and their associated measures, that are omitted from the analysis, and any tests, reasoning or special circumstances that justify their omission; and

¹³⁹ Ibid, 292-298 and 414.

- o Describe how the model specification can be interpreted to yield the measurement of program impacts.
- Error in measuring variables: Describe whether and how this issue was addressed, and what was done to minimize the problem;
- Autocorrelation: Describe any autocorrelation problems and the solutions specifically taken to address the problem. Specific identification and mitigation diagnostics should be presented, including differing treatment for sub-groups, if any;
- Heteroscedasticity: Describe the diagnostics carried out, the solutions attempted and their effects. If left untreated, explain why;
- Collinearity: Describe procedures used to address the problem of collinearity, and the reasons for either not treating it or treating it to the level that it was;
- Influential data points: Describe the influential data diagnostics that were used, and how the identified outliers were treated;
- Missing data: Describe the methods used for handling missing data during the analysis phase of the study; and
- Precision: Present the methods for the calculation of standard errors for key parameters such as gross impacts, net impacts, NTGRs, and key process and market effects measurements.

Engineering Models Including M&V

- Describe the primary sources of uncertainty in deemed and measured parameters used in engineering models;
- Describe the construction of the baseline. Include assessment and description of how the selection of baseline affects the development of gross impacts versus net impacts. Baseline definitions shall be consistent with those used in the net analysis;
- Discuss efforts to guard against measurement error associated with the various M&V data collection efforts;
- Discuss site selection and potential non-response bias, any tests performed to assess potential bias across and within site measurements, and potential effects of any remaining concerns in this area;
- Describe any potential measurement or bias issues associated with the measurement approaches and tools used as they apply to specific program parameters and estimates:
 - o Engineering model bias – systematic under- or over-prediction of effects of a measure by an engineering model;
 - o Modeler bias – the systematic under- or over-prediction of effects of a measure by a building energy simulation (e.g., DOE-2) modeler. Also includes the random under- or over-prediction of effects of a measure by a building energy simulation (e.g., DOE-2) modeler;
 - o Deemed parameter bias – systematic deviation in a deemed parameter used in an engineering model;

- o Meter bias – systematic error in meter and/or sensor;
- o Sensor placement bias – systematic over- or under-prediction of measured quantity due to sensor placement (could be combined with above); and
- o Non-random selection of equipment and/or circuits to monitor.

Summary of Sampling and Uncertainty Protocol

A summary of these Protocols is not provided here. Rather, in the summaries provided at the end of the other Protocols (Impact, M&V, Emerging Technology, Codes and Standards, Effective Useful Life and Market Effects), the relevant elements of the Sampling and Uncertainty Protocols are discussed.

Evaluation Reporting Protocol

Introduction

The Evaluation Reporting Protocol identifies the information that must be incorporated in the different types of evaluation reports and specifies how it is to be reported. This is accomplished by first identifying the common information required across all evaluation reports. Then the Protocol describes additional information and presentation formats for each of the types of evaluation reports.

The reporting information contained in this Protocol is that which support the program evaluation efforts. There are other reporting requirements associated with program status, progress and financial reporting not covered in this Protocol for which Administrators are responsible. For information relating to program status, progress and financial reporting the reader is referred to the CPUC-ED.

Report Delivery Dates

The delivery dates for each evaluation report must be identified in each program evaluation plan. Both the report delivery dates and changes to these dates must be approved by the CPUC-ED. The scheduling of the all draft and final evaluation reports must consider the timing of the information needs of the key stakeholders including the CPUC-ED, the CEC and the portfolio Administrators, so that the evaluation results can be provided in time to use the results to support program “performance basis” assessments and to support future program design and evaluation planning. This requirement does not imply that only two reports (one draft and one final) will be required from the evaluation contractor. It is expected that each evaluation will have multiple reporting periods across the multi-year study period. Each evaluation plan will detail the deliverables to be provided within the study scope and the due dates for each deliverable. Once the final reports are approved by the Joint Staff, the evaluation contractor will deliver the electronic and hard copy reports and post the final evaluation report on the CALMAC Web site consistent with the instructions detailed in this Protocol.

Common Evaluation Reporting Requirements

This section of the Reporting Protocol presents the reporting requirements specifying the information that must be reported in the various types of draft, draft-final and final evaluation reports. Typically these requirements apply to the evaluation contractors conducting the studies and preparing the reports.

The present Reporting Protocol is different than previous California reporting protocols. In addition to new evaluation reporting requirements, there are also performance basis reporting metrics that need to be reported when applicable. The evaluation contractors are responsible for knowing what information is required in their evaluation reports and for conducting the evaluation efforts in a way that provides the required information. The evaluation contractor will coordinate with the CPUC-ED to identify the performance basis reporting metrics to be included in each evaluation and structure the evaluation plan to meet those requirements. Final negotiated study-specific evaluation budgets will be structured to meet this Protocol requirement.

The following reporting requirements apply to all evaluation reports produced from studies of California's energy efficiency programs including process, impact and market effects evaluations.

The reporting requirements included in this Protocol are minimum requirements. Each program evaluation may have additional reporting requirements that are specified in the approved evaluation plan. For example, an evaluation plan may require that the evaluation report provide "designated units of measure" reporting at the program level. These units may include items such as kWh savings/square foot of commercial building served or kW savings/square foot of home served. These may also be structured so that the reporting requirements are more defined, such as kWh savings/square feet of commercial building conditioned space served, or kW savings/square foot of occupied space, heated space, cooled space, or other criteria.

1. **Draft reports are to be provided in electronic formats.** Draft and final-draft energy and load impact reports, M&V reports, codes and standards reports, emerging technology reports, effective useful life reports and draft market effects evaluation reports will be provided to the Joint Staff in electronic file formats consistent with the file format requirements provided in this Protocol for final reports (see below). Draft process evaluation reports will be provided in formats determined by the Administrators requesting the studies.
2. **At least 10 copies of all final evaluation reports must be submitted in bound hard copy format on recycled paper using double-sided printing to minimize the use of paper.** No less than four hard copies should be provided to the CPUC-ED, two hard copies to the CEC, three hard copies to the Administrator(s) for the program(s) being evaluated and one hard copy to the program implementation manager (whether a contractor or employee of the Administrator) of the program being evaluated. The Administrator and the Joint Staff can request that evaluation contractors provide additional copies as appropriate or can advise the evaluation contractors that fewer hard copies are needed. This requirement serves as the minimum deliverable of the final evaluation reports in bound hard copy format unless specified differently for an individual study.
3. **All final reports will be provided to the CPUC-ED, the CEC and the Administrators in unprotected (no password restrictions) electronic formats and protected formats that can be made available to the public.** The electronic formats must be provided in two software versions with each report provided in a single electronic file. The unprotected electronic reports must be provided in Microsoft Word®. The protected formats should be provided in Adobe® formats in a version that is loadable/readable by the organization contracting for the study. The electronic files must be named in a way that allows the recipients to understand the program or the group of programs on which the evaluation reports. Examples of acceptable file names include the following:
 - a. 06 PG&E Mass Market Process Eval.pdf
 - b. 06-08 SCE Res Programs Impact Eval.doc
 - c. 06 SCE Appliance Recycling Process Eval.pdf
 - d. 06-07 Statewide Multi-Family Programs Impact Eval.doc

Evaluation contractors conducting energy impact studies will also provide Microsoft Excel files presenting the energy savings (kW, kWh, therms) from direct or indirect impact, codes and standards, or market effects studies as described in this Protocol (see Sample Reporting Tables at the end of this Protocol).

4. **Within five days of the submission and acceptance of the final evaluation report, the organization providing the report must post it and its abstract on the CALMAC Web site using the posting instructions provided by CALMAC at the time of posting.** The abstract posted on the CALMAC site should be the one included within the final evaluation report located just after the title page. Care should be taken in developing the abstract to allow the CALMAC search engines to easily find the report when system users conduct key-word searches. Upon posting, CALMAC will distribute an e-mail announcement of the availability of the report to the CPUC's energy efficiency docket list-serve and to the CALMAC distribution list.
5. **All evaluation reports must contain the following information on the report cover of both the electronic and hard copy files.**
 - a. Report title that reflects the type(s) of evaluation(s) being conducted (e.g., Energy and Demand Impact Evaluation, Process Evaluation, Effective Useful Life Evaluation, Codes and Standards Program Evaluation, Market Effects Evaluation, or Market Effects Evaluation);
 - b. Official name of the program(s) as recorded in the CPUC's program tracking system (EEGA), including the program cycle identifier (e.g., 2006-2008, 2009-2011);
 - c. Official CPUC/EEGA tracking number(s) of the program(s) being evaluated;
 - d. Date of the evaluation report;
 - e. Name of the organization conducting the evaluation;
 - f. Name of the organization administering the evaluation;
 - g. Name of the organization administering the program; and
 - h. Name of the organization implementing the program.
6. **The title page of both hard copy and electronic formats must include the following information:**
 - a. The same information provided on the report cover, plus the following:
 - b. Name of the organization conducting the evaluation and full contact information for the evaluation lead(s) responsible for the study;
 - c. Name of the organization administering the evaluation and full contact information for the lead Administrator; and

- d. Name of the organization implementing the program and full contact information for the lead program director or manager.

(Contact information should include individual's name, address, phone number, fax number and e-mail address.)

7. **Abstract.** Following the title page, the report will include a report abstract. The abstract should be developed consistent with the "Report Summary" development instructions for posting on the CALMAC Web site. The abstract should be less than 200 words (or consistent with current CALMAC guidance) and include important key words that allow CALMAC's Web site's search engines to locate the report during routine searches.

8. **Evaluation reports should include, at a minimum, the following sections:**

- a. Cover
- b. Title Page
- c. Abstract
- d. Table of Contents
- e. Executive Summary - this section should very briefly present a review of the evaluation findings and the study's recommendations for program change, this should typically be no more than 1-3 pages. The findings and recommendations included in the summary should reference the primary text location within the report where each finding or recommendation is analyzed and presented.
- f. Introduction and Purpose of the Study - this section should give a summary overview of the evaluation and the evaluation objectives and researchable issues. This section should discuss if each of the researchable issues presented in the evaluation plan was addressed in the evaluation report and identify if any issues were not addressed and provide the reason why not.
- g. Description of Programs Covered in Study - this section should provide a description of the program(s) being evaluated in enough detail that readers can understand the program(s) and have an understanding of the program and program components that delivered the evaluation identified effects. The program description should also include the counts of the number of participants at the end of each program year for each program, and estimates of the technical potential (measure counts) for each measure covered by the program. This market potential should estimate the number of units that could be installed by the program if the technical potential was achieved for each measure covered by the program within the program's target market. The technical potential should be provided by the program Administrator and should be included in the data request delivered to the Administrators. If the Administrator does not provide the data, the report should so stipulate, identifying the data requested and the reason why the data could not be provided. If the Administrator cannot provide the requested technical potential data, the report may not be able to discuss the technical potential and the fraction of this potential achieved by the program.

- h. **Study Methodology** – this section should describe the evaluation approach in enough detail to allow a repetition of the study in a way that would produce identical or similar findings. See additional content information below.
- i. **Reliability Assessment of the Study Findings** – this should include a discussion of the threats to validity and sources of bias and the approaches used to reduce threats, reduce bias and increase the reliability of the findings, and a discussion of study findings precision levels.
- j. **Detailed Study Findings** – this section presents the study findings in detail.
- k. **Recommendations for Program Changes** - this section should be a detailed identification and discussion of the recommended changes, including the anticipated cost of the recommended change and the expected effect of the change on the operations and cost-effectiveness of the program(s).
- l. **Appendix A** - appendix A should be a presentation of the performance metrics identified by the CPUC-ED that apply to the types of programs being evaluated and a presentation of the evaluation's assessment of the performance of the program for each of the performance metrics covered in the evaluation plan.
- m. **Appendix B** - appendix B should present and discuss the success and timing of the data requests provided to the Administrators and the amount of time between the response and the receipt of the requested data. This section should discuss the success in obtaining the information needed to conduct the evaluations and identify any request made that were not provided in accordance with the provisions in this Protocol. If information was requested and not provided, the appendix should discuss the implications of not obtaining the data on the accuracy and reliability of the study findings. (Information that is maintained in the CPUC-ED program-reporting database can be obtained from the CPUC-ED and does not need to be collected from the IOUs.)

The **Study Methodology** section must include the following:

- a. Overview of the approach;
- b. Questions addressed in the evaluation;
- c. The Protocols and rigor levels assigned to the study;
- d. Description of the study methodology;
- e. How the study meets or exceeds Protocol requirements;
- f. How the study addresses issues presented in the Protocols regarding the methods;
- g. Sampling methodology;
- h. Expected precision or power analysis results (as required by the Sampling & Uncertainty Protocol);
- i. Sample descriptions (including population characteristics, contact information availability and sample disposition rates);
- j. Description of the baseline;
- k. Sources of baseline data;
- l. Description of measures; and

- m. Assumptions on measure performance (including data sources).

The **Reliability Assessment** section of the report should focus its presentation and discussion on the targeted and achieved precision levels for the key findings presented, the sources of uncertainty in the approaches used and in the key findings presented, and a discussion of how the evaluation was structured and managed to reduce or control for the sources of uncertainty. All potential threats to validity given the methodology used, as presented in the Sampling & Uncertainty Protocol, must be assessed and discussed. This section should also discuss the evaluator's opinion of how the types and levels of uncertainty affect the study findings. Findings also need to include information for estimation of required sample sizes for future evaluations and recommendations on evaluation method improvements to increase reliability, reduce or test for potential bias and increase cost efficiency in the evaluation study(ies).

The **Recommendations for Program Changes** section on need only be added when changes have been identified during the evaluation process. In general, impact evaluation studies will have the fewest program change recommendations. Market effects evaluations should provide recommendations that the evaluation contractor thinks will improve the ability of the program(s) to influence market change. Process evaluations will typically have recommendations, as generating recommendations that increase the cost-effectiveness of the program(s) is the primary purpose of conducting the process evaluation.

The evaluation reports should generally be written for a wide range of individuals, including individuals not familiar with evaluation approaches or the field's specialized terminology. Technical information needed to report methodologies used for research design, sampling, impact analysis, M&V efforts, regression and engineering analysis, bias detection, bias correction and other technical areas must be reported and should not be avoided to ensure readability by a wider range of audience. A summary of the methodology, findings and decisions covering these issues should be written for a wider audience, however the more technical details relating to these reporting categories must also be provided.

9. **Databases and analysis datasets are the property of the State of California and should be provided to the CPUC-ED within 10 working days of the acceptance of the final evaluation report.** Database and analysis datasets shall be delivered in commonly accepted formats, such as SPSS[®], SAS[®], ASCII formatted or defined fields, tab or comma delimited, ASCII text, Microsoft Excel[®], Microsoft Access[®], dBase[™] or other similarly commonly available formats. Non-common proprietary databases are not acceptable deliverable formats. Database suppliers should negotiate with the CPUC on a format structured during the evaluation planning process. Databases and analysis datasets should be provided in electronic formats with data dictionaries that describe the fields and field formats. The databases and analysis databases should be named so that they can be linked to the program being evaluated and the evaluation report presenting the findings. They should be provided so that the CPUC or their consultants can duplicate the analysis effort. If the data in the database or in the

analysis datasets is modified from the data that was collected the modifications should be disclosed.

Performance Basis Evaluation Reporting Metrics

In addition to the above-identified common reporting requirements, each evaluation should also report, in a table format, those metrics associated with the CPUC-ED's performance basis reporting requirements that are collected during the evaluation effort. While not all evaluations will collect and report all of the CPUC-ED performance basis metrics needed by the CPUC-ED, those metrics that are collected or assessed within the evaluation effort should be reported in the draft and final evaluation reports. The performance basis metrics that the evaluation should report, if collected or assessed as part of the evaluation effort are listed in Appendix C. Each evaluation contractor should identify the performance basis metrics that will be collected and assessed during the evaluation planning effort and identify those metrics that will be reported in the draft and final evaluation reports

Evaluation Type Specific Reporting Requirements

The following reporting requirements are in addition to the reporting requirements noted above and are presented for each type of evaluation and evaluation effort.

Energy Impact Evaluations and Supporting M&V Efforts

The energy impact evaluation report must focus on reporting the gross and net achieved energy savings and demand reduction that can be expected as a result of the program's efforts for each program year and for the program at the end of the program cycle in accordance with the progress of the evaluation within the program cycle. The impacts should be reported for each full calendar year (2006, 2007, 2008 and totaled for all years within a program cycle (2006-2008) over the effective useful life (EUL) of the measures installed or the behaviors changed. The reporting should assume a full year of measure use for the year in which the measures are installed. This avoids partial year reporting during the year of installation and at the end of the EUL. That is, a program that installs measures that have an effective measure life of 10 years installed in 2006, would report 10 years of savings for that measure with the first full year being 2006, regardless of the date that the measures were installed during 2006. For programs that have a mix of measures with different EULs, the savings projections will reflect the end-of-EUL drop-offs so that the projected savings represent only those savings that are expected in a specific year. When the CPUC-ED specifies that an evaluation will assess measure-level savings, the assessment should target the measures approved by the CPUC. In some cases this will encompass all of the measures included in the program and, in other cases, it will include only some specific measures.

The reported savings need also be net of interactive effects. For example, if lighting measures are installed there may be a corresponding decrease or increase in HVAC costs. Or if there are therm savings that produce an increase in electric consumption, these conditions need to be incorporated into the net effects estimate.

Savings also need to be reported by the CEC's five Climate Thermal Zones (CTZ) used for assessing Title 24 compliance, within the zones that have evaluation-study-covered program

participants.¹⁴⁰ This does not mean that M&V sampling needs to be conducted at the CTZ level, but that impact and supporting M&V results must be modeled so that the impacts are reported for each of the climate zones in which participants appear. However, the Joint Staff may request specific studies report impacts by each of the 16 climate zones in which participants appear if this requirement is in the approved evaluation plan. The CEC will provide the CTZ maps, address and geo-code matches to each climate zone and weather data to the evaluation contractors on request. Reporting also must be provided for each IOU when a program is provided in more than one IOU service territory.

Every energy impact evaluation report should include the following information:

1. **CPUC approved program ex-ante net and gross, kW, kWh and therm savings goals** recorded at the beginning of the program funding cycle and any modifications to these goals made during the funding cycle. These should be the energy savings targets for the programs included in the Administrator's portfolio filings approved by the CPUC and any changes to these goals resulting from adjustments made. If the goals have changed during the funding cycle, a brief discussion of the reasons for the change should be reported also. Goals should be reported for each calendar year in which impacts are projected.
2. **The Administrator-generated annual gross kW, kWh and therm savings.** These should be the energy and demand savings estimates that the Administrator reports to the CPUC-ED as achieved against the CPUC-approved goals.
3. **Evaluation projected annual gross and net MW (megawatt) impacts** measured for each calendar year for each year over the EUL of the measures installed or behaviors taken. Gross and net demand savings must be reported for six time periods over each of four months as follows: noon–1 p.m., 1–2 p.m., 2–3 p.m., 3–4 p.m., 4–5 p.m. and 5–6 p.m. for June, July, August and September, for each climate zone for which there are program participants. These demand savings are to be estimated using the CEC's five CTZs used for assessing Title 24 compliance. This metric represents the evaluation contractor's best estimate of the gross and net program-induced participant-based MW impacts. This metric is to be reported separately for total program savings and broken out by program-induced direct and indirect (as appropriate to each study) impacts and for participant spillover effects, if any. If the evaluation is designed to deliver measure-level kW (reported as MW) savings, the savings will be reported for each measure included in the measure-level assessment. In addition, the effects are to be reported for the measure as a whole and for both direct and indirect program effects and participant spillover effects. In addition to these reporting requirements the Joint Staff may identify additional kW reporting requirements for specific studies during the program evaluation planning process. The demand impacts are those that can be documented at the time of the evaluation and they are not to include projected impacts as a result of actions not yet taken. These impacts are not to include market effects or non-participant spillover kW effects, but instead focus only on the impacts from participants who take

¹⁴⁰ California Climate Zones, The climate zones used for this purpose are the California Energy Commission's five Climate Thermal Zones used for assessing Title-24 compliance unless specified differently by the CPUC-ED during the program planning process. In some case it may be necessary to require reporting by each of the 16 SCE Climate Thermal Zones, also referred to as the 16 Title 24 climate zones.

advantage of the program's offerings and who may replicate those actions in their facilities. If the evaluation contactor determines that MW impacts increase or degrade over time, the annual projections of impacts must incorporate that increase or degradation factor and explain the cause, the reliability and the measurement approach for documenting the increase or degradation rate. (See tables below for example of reporting formats.) Participant spillover is to be reported in the evaluation, but will not be credited for the purposes of goal accomplishment at this time.

4. **Evaluation projected annual MWh (megawatt-hours) gross and net savings** measured for each calendar year for each year over the EUL of the measures installed or behaviors taken. Savings should be reported for the program as a whole and for each of the CEC's five CTZs used for assessing Title 24 compliance in which the program operates. This metric represents the evaluation contractor's best estimate of the energy savings that will occur because of the actions of the program. There are three reporting metrics associated with this requirement. The annual MWh savings are to be reported for the program as a whole and separately, for both program participation-based direct and indirect savings, and for participant-spillover-based savings. If the evaluation is designed to deliver measure-level savings, the savings will be reported for each measure included in the measure-level assessment. The savings are those that can be documented at the time of the evaluation and they are not to include projected savings as a result of actions not yet taken. These savings are not to include market effects or non-participant spillover savings, but instead focus only on the savings from participants (direct and spillover) that take advantage of the program's offerings. If the evaluation contactor determines that savings increase or degrade over time, the annual projections of savings must incorporate that increase or degradation factor and explain the cause, the reliability and the measurement approach for documenting the increase or degradation rate. (See tables below for example of reporting formats.) Participant spillover is to be reported in the evaluation, but will not be credited for the purposes of goal accomplishment.
5. **Evaluation-projected annual gross and net therms (100,000 BTU/therm or 100 cubic feet of methane) of natural gas savings** measured for each calendar year for each year over the EUL of the measures installed or behaviors taken. This metric represents the evaluation contractor's best estimate of the energy savings that will occur because of the actions of the program. The annual therm savings are to be reported separately to include program participation-based savings plus participant spillover based savings, if any and totaled for program savings. However, participant spillover will not be counted toward the program or portfolio goal achievements. The savings are those that can be documented at the time of the evaluation and they are not to include projected savings as a result of actions not yet taken. If the evaluation is designed to estimate measure-level savings, the savings will be reported for each measure included in the measure-level assessment. These savings are not to include market effects, participant or non-participant spillover savings estimates, but instead focus only on the savings from participants that take advantage of the program's offerings. If the evaluation contactor determines that savings increase or degrade over time, the annual projections of savings must incorporate that increase or degradation factor and explain the cause, the reliability and the measurement approach for documenting the increase or degradation rate. (See tables below for example of reporting formats.)

The Energy Impacts Protocol requires the evaluation contractor to estimate annual gross and net impacts over the EUL of the installed technologies or the behavior change-induced actions. For measures like CFLs, the expected life of the impacts may only be a couple of years, while for building design changes, the impacts may be over 30 years or more if the evaluation determines that the changes would not have occurred in the absence of the program offerings. It is the responsibility of the evaluation contractor to establish evaluation designs and approaches that allow these metrics to be reported in the evaluation report. One of the primary reasons that these metrics are required is so that portfolio energy and load impact curves can be generated for each program, for each IOU and for the portfolio as a whole. These savings are not to include non-participant savings that may have been influenced by the program's operations or the spillover caused in the non-participant population as a result of the program. They are to include participant spillover or participant action replications that result as a function of program participation. However, participant spillover savings are not to be counted toward program or portfolio goal achievements.

6. **Measure counts per participant.** This metric is incorporated into the reporting criteria so that the evaluation report provides a presentation of the types of measures taken by the program participants and the number of those actions taken per participant. This metric is to be retrospective and report only the actions taken as a result of the program at the time of the evaluation. However, the evaluation should true up these metrics at the end of each program year so that they can be reported for each program year (see sample reporting sheet at the end of this Protocol). The assessment should be based upon tracking system reviews and informed by the impact evaluation and the supportive M&V efforts. It can also be supported by the process evaluation efforts, if there is coordination among the evaluation efforts. The evaluation study can also separately report projected actions to be taken, if approved in the evaluation plan.
7. **Measure counts versus program goals.** This metric is incorporated into the reporting criteria so that the evaluation report provides a presentation of the evaluation verified program accomplishments relative to measure installation goals. This metric is to be retrospective and report only the actions taken as a result of the program at the time of the evaluation. It should be based upon tracking system reviews and informed by the impact evaluation and the supportive M&V efforts. It can also be supported by the process evaluation efforts, if there is coordination among the evaluation efforts. The evaluation study can also report projected actions to be taken, if approved in the evaluation plan.
8. **Measure-level savings.** If the evaluation plan is structured to provide measure-level or behavior-level savings estimates, these metrics should be reported for the covered measures. Not all program evaluation plans will be focused at the measure or behavior level. However, for those that are, as a result of the Joint Staff evaluation prioritization efforts, the savings should be reported at the measure or behavior level. In these cases, the program evaluation report should specify the program offering and design conditions that lead to the measure-level savings and the measure use conditions that affect the savings. The purpose of this requirement is to be able to update the DEER database estimates by changing current estimates, as new data are developed and add new measure classifications to the DEER

database when it is apparent that the program design and operational conditions affect the level of energy and demand savings.

9. **Measurement reliability metrics.** Results and all measurement reliability information must be reported at the program level, program group level and for any program component or delivery mechanism with a designated separate level of rigor or as designed in the approved evaluation plan. In addition, the following data reliability metrics should be reported for the energy impact estimates provided in the evaluation report.

- a. Precision level at the 90% confidence level of the direct participation energy savings (kWh/MWh);
- b. Precision level at the 90% confidence level of the participant spillover energy impacts (kWh/MWh) (if available separately given the methodology selected);
- c. Precision level at the 90% confidence level of the direct demand energy impact (kW/MW);
- d. Precision level at the 90% confidence level of the participant spillover demand impacts (kW/MW) (if available separately given the methodology selected);
- e. Coefficient of variation (CV) or standard deviations (SD) and means on the realization rate(s) for the program's energy effects and for all strata in any stratified sampling effort; and
- f. P values for all energy impact estimates (kW, kWh, therms).

10. **Savings comparison.** The report should include a presentation and discussion of the CPUC approved program goals compared to the estimated realized savings from the evaluation findings (this should be expanded in the Appendix described below).

11. **Appendix C.** Appendix C should present, assess and discuss the similarities and differences between Administrator savings assumptions and projections, and the results of the evaluation findings. This discussion should identify what assumptions were confirmed and not confirmed, and identify recommended changes to the assumptions that Administrators use to project savings.

12. **Appendix D.** Appendix D should present the weather data used to conduct the evaluation, including the heating and cooling degree-days used in the study, if any.

Note: See end of Reporting Protocol chapter for examples of energy impact reporting tables.

Measurement and Verification (M&V)

1. **M&V plan and reporting requirements.** For impact evaluations that are supported by measurement and verification (M&V) efforts, the evaluation report should present the program-specific M&V plan in enough detail that the plan can be replicated. The plan should describe and/or discuss:

- a. How the M&V samples were identified and selected;

- b. How the M&V activities were used to support the impact assessment;
- c. Any disagreement between the sampling plan and the sampling approach used, and how the difference influences the reliability of the study findings;
- d. Sampling and measurement bias issues and how these biases can be expected to influence the impact estimates;
- e. How the biases were controlled or mitigated in the M&V efforts and what statistical or measurement approaches were used to adjust the M&V data to inform the impact estimates; and
- f. How the M&V results were used to estimate net program energy impacts.

Justification for the identification and selection of the baseline is required. An assessment and discussion of the baseline selected and its consistency of use for gross and net impacts must be included.

Site-specific M&V plans prepared during the course of the study shall be provided in an Appendix to the impact evaluation report. The site-specific M&V plan shall include all topics specified in the M&V Protocol, including assumptions used for stipulated parameters, the source of the assumptions and uncertainties associated with the M&V study results. Measure-level M&V results shall be reported according to the applicable DEER-compatible format listing in Appendix A of the M&V Protocol. Energy and peak demand savings resulting from weather dependent measures shall be reported under weather conditions prevailing during the course of the M&V project. These weather conditions shall be reported along with the energy and peak demand impact information. The impacts shall be normalized to standard weather conditions consistent with the CEC CTZ long-term average weather conditions for the climate zone in which the site is located.

2. **M&V analysis database.** The M&V analysis database(s) will be provided to the CPUC-ED upon delivery of the evaluation report. Site-specific M&V results shall be reported electronically according to database formats established by the Joint Staff compatible with EEGA and DEER databases. Field data shall be supplied in a non-proprietary format, such as ASCII text, Microsoft Excel, Microsoft Access, dBase or XML for inclusion in an M&V data warehouse. Proprietary databases are not acceptable deliverable formats. Building characteristics data collected during on-site surveys shall be reported according to the International Alliance for Interoperability (IAI) Industry Foundation Class (IFC), ifcXML or aecXML formats; Green Building XML (gbXML) format or other electronic data formats as designated by the Joint Staff. The Joint Staff will establish procedures to submit, receive and store M&V database(s) within a data repository. Because the databases will, by their very nature consist of customer-specific information, they will be secured and safeguarded against public release. Evaluation contractors will be informed of these instructions during the evaluation planning process.

Emerging Technology Program Evaluations

The Emerging Technology Program Evaluation will be reported consistent with the requirements for all reports described in this Protocol under Common Evaluation Reporting Requirements (above). In addition, the following elements should be included in the evaluation reports under the Methods heading.

- Program Theory and Logic Model
- Goal Verification
- Aggregate-Level Analysis
- Implementation Analysis
- Measure Tracking
- Detailed Analysis of Key Performance Indicators
- Peer Review
- Target Audience Surveys

These presentations must be provided in enough detail that the differences (if any) in the methodological approach across different technologies and utilities can be understood by the reader. Finally, one must describe the approach for integrating the study results so that the overall performance of the ETP can be assessed.

The Reporting Protocols includes a requirement that all evaluation reports include a presentation of the detailed study findings. This presentation must be provided in enough detail that the different results or findings (if any) can be understood for each technology assessment covered in the study. The report should present the results of each of the required eight components contained in the ETP Protocol. Reports will be provided consistent with the Reporting Protocol.

Codes and Standards Program Evaluations

The Codes and Standards Program Evaluation will be reported consistent with the requirements for all reports described above (Common Evaluation Reporting Requirements) and shall also present the following information.

1. **Change theories.** The report should present each of the code or standard change theories in an appendix to allow the reader to understand the theory behind the change achieved. The report should include a brief summary of the change theories in the text of the report.
2. **Change timelines.** The report should present a timeline associated with the program's efforts employed to influence changes for each code or standard change influenced by the program. The timeline should begin with the time at which the code targeting and selection effort was launched and end with the code adoption date. The code adoption date should be followed by the date that the change takes effect.
3. **Overview of the program activities that caused the change.** The report should provide a discussion of the activities and events that are wholly or in part responsible for the program-induced changes.
4. **Summary code or standard changes.** The report should present a summary of the change to the code or standard caused by the actions of the program. The summary should be detailed enough for the reader to understand the change that occurred and the significance of the change to the level of savings predicted.
5. **Jurisdictions.** The report should discuss the jurisdictions covered and not covered by the changes for each change included in the study.

6. **Listing of meetings, events, activities and documents.** The report should present a listing of each of the meetings, events, and activities attended or monitored by the evaluation contractor to support the evaluation effort and a listing of the documents reviewed to support the study.
7. **Interviewees.** The report should provided the titles of all interviewees providing information used in the analysis. The contractor should report the names of the interviewees in an accompanying memorandum, but not place the names of these individuals in the final public document. Note: individual interview results should be treated as confidential information.
8. **Pre-change penetration rates.** The report should present the pre-change technology adoption or penetration rates reported by the program.
9. **Naturally occurring market adoption.** The report should describe the approach for estimating the naturally occurring code and standard changes and the results of applying the approach for each change.
10. **Attribution approach and results.** The report should discuss the program attribution analysis approach and results of that approach for each change covered in the study.
11. **Gross market-level energy impacts.** The report should present the approach for estimating the gross market-level energy impacts for each change and provide the results of that analysis.
12. **Net market-level energy impacts.** The reports should present the net energy impact adjustment approach and the results of that approach so that the reader can understand the influence of each adjustment on the resulting net savings. The contractor should report net savings for each change and report the resulting net savings on the reporting spreadsheets (see Sample Reporting Tables).

Market Effects Evaluations

The reporting for the market effects evaluation must include the following information.

1. **Market theory integrated with program theory.** The report should clearly present and describe the market theory and, if constructed, the market logic model. The program and market theory should be integrated so that the anticipated net market effects (those market effects induced by program interventions) can be more readily perceived within the context of the market theory. This will provide a more comprehensive framework from which to conduct the market effects evaluation. The market theory and program theory(ies) should provide the following information:
 - a. The **market theory** should be described in detail, including how the markets operates, its structure and scope, and how the various energy efficiency interventions are expected to change the market. The market theory should include how other market actors, activities and interventions are functioning to change the market that may work in sync or in opposition to the energy efficiency programs. The market theory should present a comprehensive view of how the market operates and address how, when, where, why and under what conditions market effects are expected to occur. The report should identify

the specific changes in the market that can be observed and measured if the market is being changed. The market theory should describe the individual questions that were asked and the indicators or metrics monitored to assess when, how and to what degree the market is being changed and it should identify key market conditions that influence market change and the rates of change.

b. Program theory. If one or more of the programs expected to cause changes to the operations of a market is designed specifically to cause a market effect (change in the operations of a market defined for the evaluation), the report should present and describe the program theory(ies). The evaluation report should present and discuss the program theory used by program managers to structure their change efforts and explain how the program theory was used to focus the evaluation efforts. If a program logic model is developed it should be presented. The program theory should present how the program's operations lead to observable and measurable market effects. It should show the resources placed into the market, how they are placed, the resulting planned activities and the anticipated outcomes (the end effects) from those activities. The program theory should identify the key market metrics or measurement points that are expected to change as a result of the program's efforts and actions.

2. **Assessment of gross measure or behavior change.** The report should present and describe how the gross level of measure and/or behavior change in the market is being measured and confirmed through data collection, change measurements or change verification efforts. The report should identify any primary or secondary data used to estimate baseline or current condition measure or behavior use status. The report should describe how both baseline market conditions and current conditions are quantified and how gross measure or behavior-use conditions are being estimated. The report should discuss the reliability of these estimation methods and the various threats to the validity and accuracy of the estimation approach. The report should present the results of a Monte Carlo or other risk assessment approach that examines the difference in report conclusions that would occur if the key assumptions in the establishment of the gross measure or behavior change vary within reasonable levels of variance. This activity should result in the assignment of gross measure use or behavior change conditions that have resulted in the market as a result of all market effects, including program induced and non-program induced effects, and the presentation of the degree of variance that could be expected within those measurement conditions.
3. **Gross and net market change attribution assignments.** The report should explain the rationale behind the study's approach for identifying and allocating causal actions and activities across the market change metrics and change indicators and identify how net program-induced market change will be identified. The report should present the sources of market change and describe how allocation of the cause of the change is being proportioned across the various change agents (program and non-program influenced), so that all observed changes in the market are assigned to one or more reasons for the observed change. The report should present and discuss the proportioning approach and any data weighting or assignment systems used, and justify why the assignment approach is reliable and representative of how the market works. The report should discuss any inconsistencies between the allocation approach and the market theory discussed earlier.

4. **Net market change.** The report should present the results of applying the attribution assessment with the gross measure or behavior change assessment (previous two reporting activities) to identify the net program-induced measure or behavior changes and to identify the programs or program events that are included in the program-related attribution assignments. This assessment will identify the proportion of the market change that is caused by the program(s).
5. **Assignment of energy impacts.** The report is to present the results of the assignment of energy impacts resulting from applying energy and demand impacts associated with the net measure changes in the market that are caused by the program's efforts. The accepted practice for this assignment is to use the energy and demand savings for the covered technologies or behaviors reported in the latest DEER update. For measures not included in the DEER update, the evaluation will report the best engineering assessment estimation approach for the technologies or behaviors not included in the DEER database as guided by the results of the most recent evaluations of those technologies, if any, with appropriate references and justification for their applicability to this analysis. If a net effects modeling approach is used the steps of the process must be clearly described in a manner consistent to permit the analysis to be repeated by other researchers.

Process Evaluations

The process evaluation report shall include the following reporting requirements in addition to the common evaluation reporting requirements presented earlier:

1. **Detailed program description.** While all evaluation reports are to have a description of the program(s) covered in the evaluations, the process evaluation report must present a detailed operational description of the program that focuses on the program components being evaluated. Use of a program flow model is highly recommended. The reader of the report must be able to understand the operations of the program being evaluated in significant enough detail that they can understand the components of the program that would be affected by the program change recommendations.
2. **Program theory.** The process evaluation should include a presentation of the program theory. The program theory should, when possible, be the theory developed or approved by the Administrators. If the Administrators have not developed a program theory, they should be provided with the opportunity to develop the theory for inclusion in the evaluation report. If the detailed program theory is not available or cannot be provided in time for the evaluation report due date, the evaluator should include a summary program theory built from the evaluation team's program knowledge. This theory does not have to be approved or reviewed by the Administrator to be included in the evaluation report. The purpose of this requirement is to have a complete program description in the evaluation report and provide the Administrators the opportunity to provide the included program theory, but not to burden the Administrator with the development of the program theory or logic models if they have not already been developed. If the evaluation contractor develops the program theory or the associated logic model, it should be noted as such and complete enough for the reader to

understand the environment in which the program recommendations are to be placed, but does not need to be a finely detailed program theory or logic model.

3. **Support for recommended program changes.** While all evaluation reports are expected to have a section on recommended program changes, identifying these recommendations is one of the primary purposes of the process evaluation report. All recommendations need to be adequately supported, per the Protocol requirements. Each recommendation should be included in the Executive Summary and then presented in the Findings text along with the analysis conducted and the theoretical basis for making the recommendation. The Findings section should include a description on how the recommendation is expected to help the program, including the expected effect implementing the change will have on the operations of the program. The Findings section should include a discussion on how the recommended change can be made, who should be responsible for making the change and the expected cost and benefits of the change. If the information to conduct a cost-benefit forecast/prediction for the recommended changes is collected as part of the approved evaluation plan, the report should include a cost-benefit assessment of the recommendation so that the cost of the change can be compared to the expected benefits.
4. **Detailed presentation of findings.** A detailed presentation of the findings from the study is required. The Findings should convey the conditions of the program being evaluated and should be presented in enough detail that any reader can understand them and the associated implications to the cost-effective operations of the program. (See 3 above for more details on content requirements of the Findings section.)

Effective Useful Life Evaluations

The Effective Useful Life Evaluation will be reported consistent with the requirements for all reports described above (Common Evaluation Reporting Requirements) and shall also present the following information.

All EUL evaluations are expected to assess and discuss the differences between (a) the ex-ante EUL estimates from DEER or as otherwise approved by the Joint Staff and (b) the ex-post EUL estimates produced by the EUL evaluation study(ies). To the extent that the data gathered and evaluation analyses conducted can explain the causes for these differences, this must be presented and discussed. The evaluation report should note situations in which explanations are not possible due to lack of sufficient data or problems with interpretation. The EUL evaluation report must also include a recommendation of the EUL for the measure and delivery strategy/application that should be used for future program planning. This recommendation may take the form of recommending the replacement of a DEER EUL or the establishment of a new DEER category.

The EUL studies are also required to report the findings from the most recent degradation studies related to the EUL study measures/applications, so that the EUL report is a depository for all current persistence studies for the study measures/applications. This will assist Joint Staff and future evaluators in finding all relevant persistence information for a measure/application in one location.

All reporting under the **Effective Useful Life** Protocol should include the following:

11. Cover page containing the measures and delivery strategies or applications included in the retention evaluation, program names in the portfolios over the last 5 years that include these, program administrators for these programs and their program tracking number(s), evaluation contractor, and the date of evaluation plan.
12. Table of Contents.
13. High-level summary overview of the measures and delivery strategies or applications included in the evaluation, the programs affected, and the evaluation efforts.
14. Presentation of the evaluation goals and researchable issues addressed in the evaluation.
15. Description of how the evaluation addresses the researchable issues, including a description of the evaluation priorities and the use of assigned rigor levels to address these priorities.
16. Detailed description of the data collection and analysis methodology.
17. Current and prior retention results for selected measures given delivery strategy/application and their precision levels at a 90% confidence interval.
18. Retention, degradation, and EUL findings as is appropriate for the study assigned.
19. A discussion of the reliability assessment to be conducted, including a discussion of the expected threats to validity, sources of bias, and a short description of the approaches planned to reduce threats, bias, and uncertainty.
20. Contact information for the evaluation manager, including address, telephone numbers, fax number and e-mail address.

In addition to the above requirements, **retention** studies must also include the following:

2. Description of initial and final sample of measures still surviving.
3. Description of any findings on factors leading to the higher or lower retention rates.
4. Description of removal reasons, their distribution, and potential issues created by different removal reasons and the research design and functional forms that should be investigated in future EUL studies for these measures.

In addition to the overall EUL study reporting requirements, **degradation studies** must also include the following:

1. Describe any findings on factors leading to the relative degradation rates and absolute degradation rates, if available.
2. Describe the impact of degradation on energy savings

In addition to the overall EUL study reporting requirements, **EUL analysis studies** must also include the following:

1. Specific equations for survival functions and estimated precision of curve fit.

2. Analysis of the ex-post EUL compared to the ex-ante EUL and comparison of to the methods and results from any prior retention, degradation, or EUL studies available for that measure (to include comparisons by delivery strategy and application).
3. Recommended EUL for the measure and delivery strategy/application that should be used for future program planning.

Additional Information

1. All evaluation reports are public property and are owned by the State of California.
2. No reports or other deliverables will contain information that allows examiners of the final delivered reports and databases to be able to identify individual residential or non-residential customers or their energy consumption, energy demand or energy costs. Individual customer or participant information is to be treated as confidential information and protected by confidentiality agreements. Customer-specific information will be safeguarded from public access.
3. Customer information developed in the evaluation efforts or used to support the evaluation efforts will be maintained for a limited period of time consistent with the needs of the evaluation efforts and to support time-series or time-sensitive analysis, but will not be indefinitely maintained. All customer-specific information will be maintained and protected from disclosure for as long as there is an evaluation plan covering the use of the data to support an evaluation effort. Once there is no evaluation plan covering the use of the customer-specific data, it will be deleted or discarded in the following ways:
 - a. Electronic data files will be deleted from electronic storage systems;
 - b. Hard copy data files will be shredded and recycled or discarded if they cannot be recycled;
 - c. Electronic data file medium (e.g., DVDs, CD, electronic tape) will be shredded and recycled or discarded if it cannot be recycled; and
 - d. Other materials containing customer-specific information will be rendered unreadable and recycled or discarded.
4. The CPUC-ED will develop a data archive plan for the housing, maintenance, supervision and protection of evaluation-related data. The data archival plan shall support the segregation of data by the type of the data stored (e.g., program participation data, market description information, customer metered data and evaluation analysis data). These examples are provided to be exemplary only and are not intended to identify or define the categories within the archive plan.

Sample Reporting Tables

The following section of the Protocols provides the reporting tables that are to be completed by the evaluation contractors and provided in hard copy format in an appendix, and electronic format with the evaluation reports. A more complete set of tables will be provided by the Joint Staff in Microsoft Excel formats for evaluation contractors to use to report study results

following the approval and distribution of the Reporting Protocol. Energy impacts should be reported for each program. In addition, energy impacts should be reported as specified in each approved evaluation plan. In some cases this will require energy impact reporting by delivery strategy or other approach. It is important that the evaluation study reports the total savings for the programs being evaluated so that full credit for the energy savings impacts can be recorded for each program.

Example Megawatt Reporting Table (1) – Program Wide MW savings across all climate thermal zones

This table reports the total program participation megawatts saved across all climate zones in which the program was offered. If the program covers more than one IOU territory, tables for each IOU should also be reported. Separate MW savings tables should be prepared for the months of June, July, August and September. See Reporting Protocol for additional information. Table reports the Administrator-forecasted MW savings (gross and net), the Administrator-reported MW savings (gross and net) and the evaluation-reported gross, net of free-riders and spillover MW savings.

Total Program MW Savings																
PG&E Residential Direct Install Widget Program						June of 2006 Evaluation Projected Demand Impacts (MW average weekday across periods)										
Program ID #: 1234-06																
All Program Measures																
Climate Zone: All Program Covered Climate Zones																
Savings Timeline		Administrator forecasted ex- ante MW (CPUC approved)	Administrator forecasted ex- ante MW (CPUC approved)	Administrator reported ex- ante MW	Administrator reported ex- anti MW											
Program Year	Calendar Year	Gross	Net of Freeriders	Gross	Net of Freeriders	Noon-1PM			1PM-2PM			2PM-3PM			3PM-4PM	
						Gross	Net of Freeriders	Participant Spillover	Gross	Net of Freeriders	Participant Spillover	Gross	Net of Freeriders	Participant Spillover	Gross	Net of Freeriders
1	2006															
2	2007															
3	2008															
4	2009															
5	2010															
6	2011															
7	2012															
8	2013															
9	2014															
10	2015															
11	2016															
12	2017															
13	2018															
14	2019															
15	2020															
16	2021															
17	2022															
18	2023															
19	2024															
20	2025															
Add more rows if approved in evaluation plan																

Add more rows if approved in evaluation plan

Evaluators' Protocols

Example MW Reporting Table (2) – Program MW savings for a specific climate zone.

This table provides an example of a CEC thermal climate zone (CTZ)-specific MW savings reporting table. (CEC CTZ-1).

Total Program MW Savings																	
PG&E Residential Direct Install Widget Program						June of 2006 Evaluation Projected Demand Impacts (MW average weekday across periods)											
Program ID #: 1234-06																	
All Program Measures																	
Climate Zone: CEC CTZ 1																	
Savings Timeline		Administrator forecasted ex- ante MW (CPUC approved)	Administrator forecasted ex- ante MW (CPUC approved)	Administrator reported ex- ante MW	Administrator reported ex- anti MW												
Program Year	Calendar Year	Gross	Net of Freeriders	Gross	Net of Freeriders	Noon-1PM			1PM-2PM			2PM-3PM			3PM-4PM		
						Gross	Net of Freeriders	Participant Spillover	Gross	Net of Freeriders	Participant Spillover	Gross	Net of Freeriders	Participant Spillover	Gross	Net of Freeriders	Participant Spillover
1	2006																
2	2007																
3	2008																
4	2009																
5	2010																
6	2011																
7	2012																
8	2013																
9	2014																
10	2015																
11	2016																
12	2017																
13	2018																
14	2019																
15	2020																
16	2021																
17	2022																
18	2023																
19	2024																
20	2025																
Add more rows if approved in evaluation plan																	

The savings reported in this table are only those that occur within thermal climate zone (CTZ) 1. These tables are for reporting program MW impacts in specific CTZs for programs offered and with participants in more than one CTZ.

Example MW Reporting Table (3) – Measure-specific, program wide MW – all climate zones

This table provides an example of the reporting requirements for individual measures assessed in the evaluation study if the evaluation plan is approved to address measure assessments. It presents the measure-specific program savings for “*Measure X*” for all climate zones. In this example the measure-specific impacts are provided at the program level, not at the climate zone level. In some cases, if the evaluation plan specifies it, impacts may be required to be reported at the climate zone level.

Measure Specific MW Savings																	
PG&E Residential Direct Install Widget Program						June of 2006 Evaluation Projected Demand Impacts (MW average weekday across periods)											
Program ID #: 1234-06																	
Measure X																	
Climate Zone: All Program Covered Climate Zones																	
Savings Timeline		Administrator forecasted ex-ante MW (CPUC approved)	Administrator forecasted ex-ante MW (CPUC approved)	Administrator reported ex- ante MW	Administrator reported ex- anti MW												
						Noon-1PM			1PM-2PM			2PM-3PM			3PM-4PM		
Program Year	Calendar Year	Gross	Net of Freeriders	Gross	Net of Freeriders	Gross	Net of Freeriders	Participant Spillover	Gross	Net of Freeriders	Participant Spillover	Gross	Net of Freeriders	Participant Spillover	Gross	Net of Freeriders	Participant Spillover
1	2006																
2	2007																
3	2008																
4	2009																
5	2010																
6	2011																
7	2012																
8	2013																
9	2014																
10	2015																
11	2016																
12	2017																
13	2018																
14	2019																
15	2020																
16	2021																
17	2022																
18	2023																
19	2024																
20	2025																
Add more rows if approved in evaluation plan																	

Example Megawatt hours Reporting Table (4) – Program Wide annual MWh impacts

This table provides an example of the table to be used for reporting program wide MWh savings. In this case the table is for all program measures across all CEC CTZ covered by the sample program.

Total Program MWh Savings										
PG&E Residential Direct Install Widget Program			Administrator forecasted ex-ante MWh savings (CPUC approved)	Administrator forecasted ex-ante MWh savings (CPUC approved)	Administrator reported ex-ante MWh savings	Administrator reported ex-anti MWh savings	Evaluation Estimated Annual MWh savings	Evaluation Estimated Annual MWh savings	Evaluation Estimated Annual MWh savings	Evaluation Estimated Annual MWh savings
Program ID#: 1234-06										
All Program Measures										
Climate Zone: All Program Covered Climate Zones										
Savings Timeline			Gross	Net of Freeriders	Goss	Net of Freeriders	Gross	Freeriders	Participant Spillover	Net of Freeriders
Program Year	Calendar Year									
1	2006									
2	2007									
3	2008									
4	2009									
5	2010									
6	2011									
7	2012									
8	2013									
9	2014									
10	2015									
11	2016									
12	2017									
13	2018									
14	2019									
15	2020									
16	2021									
17	2022									
18	2023									
19	2024									
20	2025									
Add more rows if approved in the evaluation plan.										

Example Natural Gas Reporting Table (5) – Program wide annual Therms of natural gas savings

This sample table is used to report the total program energy savings in therms of natural gas for a sample program.

Total Program Therm Savings									
PG&E Residential Direct Install Widget Program		Administrator forecasted ex- ante therm savings (CPUC approved)	Administrator forecasted ex- ante therm savings (CPUC approved)	Administrator reported ex-ante therm savings	Administrator reported ex-anti therm savings	Evaluation estimated therm savings	Evaluation estimated therm savings	Evaluation Estimated therm savings	Evaluation Estimated therm savings
Program ID#: 1234-06									
All Program Measures									
Climate Zone: All Program Covered Climate Zones									
Savings Timeline		Gross	Net of Freeriders	Goss	Net of Freeriders	Gross	Freeriders	Participant Spillover	Net of Freeriders
Program Year	Calendar Year								
1	2006								
2	2007								
3	2008								
4	2009								
5	2010								
6	2011								
7	2012								
8	2013								
9	2014								
10	2015								
11	2016								
12	2017								
13	2018								
14	2019								
15	2020								
16	2021								
17	2022								
18	2023								
19	2024								
20	2025								
Add more rows if approved in the evaluation plan.									

Example Measure Count Reporting Table (6) – Units installed by program year

This table provides an example of the table to be used to report the number of measures the Administrator projects to be installed over the program cycle and the number of measures estimated via evaluation efforts to be installed. As the evaluation contractor conducts the evaluation, the results of interviews, surveys, monitoring and verification efforts will allow the contractor to estimate the number of units to be installed for each of the measures offered through the program. This table is to be provided in each of the impact evaluation reports to the extent it can be completed at the time of the evaluation effort. In addition, after the last year of the program cycle, the evaluation contractor is to complete this table for each year in the program cycle, summing the total evaluation estimated installs for the program cycle and assess the difference between the Administrator-projected installs and the evaluation-estimated installs.

Measure Counts Reporting								
PG&E Residential Direct Install Widget Program	CPUC-ED Approved Units of Measure							
Program ID#: 1234-05	Total 2006-2008 program administrator projected units to be installed	2006 Evaluation estimated total units installed	2006 Evaluation estimated average units per participant	2007 Evaluation estimated total units installed	2007 Evaluation estimated average units per participant	2008 Evaluation estimated total units installed	2008 Evaluation estimated average units per participant	Total 2006- 2008 Evaluation estimated units installed
(use CPUC approved standard measure descriptions)								
Measure A								
Measure B								
Measure C								
Measure D								
Measure E								
Measure F								
Measure G								
Measure H								
Measure I								
Measure J								
Measure K								
Measure L								
Measure M								
Measure N								
Etc.								
Etc.								
Add more lines if needed.								

The above tables are examples of the energy impact reporting tables to be provided from the evaluation efforts. A final, more comprehensive set of tables will be developed by the Joint Staff and distributed by the CPUC-ED once the Protocols are approved. The tables will be developed in Microsoft Excel so that they can be populated by the evaluation contractors and reported in the evaluation reports and delivered as separate Excel files.

Evaluation Support Information Needed From Administrators

This section of the Protocols presents the types of information that the Administrators will need to provide in support of the evaluation efforts. Not all of the information listed below will be needed for all evaluations. Rather, each evaluation will need a somewhat different set of information from the types of information presented in this chapter. In some specific cases the evaluation contractor may need to request information that is not detailed below in order to conduct an evaluation. As noted in the above individual Protocols, each evaluation plan will describe the type of information that the evaluation contractors need to complete their study. This allows the Administrators to have an advanced notice of the type of information that will be requested via a formal data request and an understanding of how the data is to be used in the evaluation study. In requesting data from the Administrators the evaluation contractors should first determine if the data needed is available in the CPUC-ED's program tracking and reporting database. If the required information is available from the CPUC-ED's database, evaluation contractors should obtain it directly from the CPUC-ED.

The information needed from the Administrators is considered basic program and participant tracking information. However the Joint Staff realizes that there may be circumstances when the requested information is not available from the Administrators or may be not be available in electronic formats. If an Administrator is unable to provide the requested information for a specific program, the Administrator will advise the organization requesting the information and the Joint Staff that the information is not available and explain the reasons why. If the requested information is available only in hard copy records, the Administrator will inform the requesting evaluation contractor and the Joint Staff and they will come to an agreement on the what information should be provided and in what format.

It is expected that the Administrators will respond to all evaluation data requests within 30 working days by providing as much of the requested information covered in this Reporting Protocol as possible, in formats agreed upon by the Administrators and the evaluation team leads. It is expected that information not covered in this Reporting Protocol, but that is necessary to conduct the evaluation in accordance with approved evaluation plans, will also be provided within 30 working days of the receipt of a data request. If this timeline cannot be met by the Administrators, the Administrators will provide the requesting organization and the CPUC-ED with an explanation of why the timeline cannot be met and work with the CPUC-ED and the evaluation contractor to establish a mutually agreed upon delivery timeline. The Administrator will inform the CPUC-ED's evaluation manager when the requested data has been provided to the evaluation contractor.

It is the Administrator's responsibility to establish and maintain program-tracking systems that are capable of supporting the evaluation efforts and of meeting the requirements specified in this Protocol. Joint Staff

All evaluation-related data requests will be provided to the appropriate Administrator(s) and the CPUC-ED at the same time. No evaluation contractor will contact customers or participants for

evaluation information without the approval of the CPUC-ED and at least 15 days advanced notification provided to the appropriate Administrator(s).

The Joint Staff understands that in some cases the data requested may not be available from the Administrator and it will be the responsibility of the Administrator, the CPUC-ED, the CEC or the evaluation contractor to collect the needed data. When the needed data is not available from the Administrators, the evaluation plan should address how the data will be collected. Again, the evaluation contractors should consider the information available in the CPUC-ED's program tracking database to determine what data needs to be collected from the Administrators and what data can be collected directly from the CPUC-ED's program tracking database.

The evaluation contractor will limit all data requests to information critical to the success of the evaluation. Information requests will be for enough data to successfully conduct the study at the needed population sample sizes, but will not over-request sample points beyond what is needed to conduct the evaluation. Evaluation data requests will need to plan for sample erosion due to a wide variety of conditions.

All measure information must be reported by the Administrators and the evaluation contractor by the CPUC-ED-approved measure description list so that identical measures are described using the same terms and definitions. If no list exists at the time of this Protocol, the Joint Staff will develop a uniform measure description list that all parties will use (IOUs, contractors, third-party providers, CPUC-ED, CEC and other stakeholders) and distribute that list via the CPUC-ED's web page and the energy efficiency list serve. The descriptions will include an official identifier and an abbreviated term that can be used in tracking systems, tables and charts where space is limited.

Most Administrators will find that the reporting requirements are consistent with the type of information that evaluation contractors have requested in the past. However, there are some examples of information detailed below that may not be routinely maintained for each program or updated on a regular basis. As a result, there may be some specific parts of a data request that the Administrators will be unable to provide. In general, the Administrators are responsible for providing the Administrator-collected or implementer-collected information requested by the evaluation contractors to allow the evaluations to be conducted consistent with the evaluation plans approved by the Joint Staff or by the CPUC-ED.

The following data should be readily available from the Administrators.

Program Information

1. Full program descriptions, including operational or procedures manuals and activities descriptions and description of implementation territories;
2. Detailed descriptions of tracking system and tracking system operations, including data dictionaries;
3. Program management and staff names, titles, work locations, phone numbers, fax numbers, e-mail addresses;

4. Program theories and associated logic models if developed. If not developed a statement that they have not been developed with a projected date of delivery of the completed theories and logic models;
5. Market operations theories describing the operations of the markets in which the program operates and, if available, a description of how the program is to change the operations of the market;
6. A description of the size of the market targeted by the program, and a description of the baseline conditions at the measure/behavior level and a discussion of how the program is expected to change baseline measure/behavior conditions, if available;
7. A description of the pre-program technical potential at the measure/behavior level and a projection of the remaining technical potential at the end of the program cycle, if available; and
8. When the program relies on key market actors, trade allies and other stakeholders to deliver or support the program in order to reach the energy saving or outreach goals, the Administrator should provide a listing, description of and contact information for these individuals/organizations.

Participant Data

For the purposes of this Protocol a participant is defined as an individual or an organization that receives a program service or financial incentive. For most programs, participants are clearly defined in the program tracking systems. However, there are times when a participant is not clearly defined or is not easily identified. The CPUC-ED expects that the Administrators will focus efforts on collecting participant information to the extent possible and practical for various types of programs or program services. Participants in resource programs are generally easy to identify as they directly receive a service or a financial incentive. Participants in other programs, such as marketing and outreach programs can be harder to identify and report. This Protocol does not act to require all programs to identify all participants. However when participant information is collected by the Administrators or their subcontractor, much of this information will be of value to the evaluation efforts. It is to the responsibility of the Administrators to work with their subcontractors to assure that when possible and practical the following information should be collected and maintained.

The following participant data should be available in electronic form with supporting database dictionaries to the evaluation teams on request.

Non-residential program data requests for end-user focused programs

1. Name of program(s) or program component(s);
2. Name of firms participating in program or program component;
3. Service turn on date;
4. Primary and secondary NAIC codes associated with the participants if available;
5. Extent to which customer is a repeat participant or a participant in other programs over the previous five years, if available or accessible;
6. Pre-participation measure and measure-use information, descriptions and conditions;
7. Address(es) of the participating firms or key participation decision makers;

8. Address(es) where program-related action is taken or for the services received;
9. Listing or description of actions taken or services received for each location by measure and end-use according to standard measure and end-use definitions established herein. These lists and descriptions should, to the extent possible, be standardized so that all Administrators use the same term for the same measure;
10. Individual participation contact information for each location to include:
 - a. First and last name;
 - b. Address;
 - c. Phone number;
 - d. Fax number (if collected); and
 - e. E-mail address (if collected).
11. Dates of key action/activity/installation steps associated with program participation:
 - a. Program enrollment date(s);
 - b. Rebate or incentive payment date(s);
 - c. Measure install dates;
 - d. Date of training received; and
 - e. Post-installation measure inspection dates.
12. Financial assistance amounts paid to participant by measure or action taken;
13. Project description information;
14. Estimated savings for actions taken;
15. Summary characteristics of building on which actions are taken or the operational environment in which measures are installed if collected;
16. Account and meter numbers and consumption histories from utility bills from all relevant meters for at least twelve months prior to program enrollment date and through to current period. Note: The evaluation contractor will work with the IOUs to understand what metered data is available for which types of customers and the formats and time intervals associated with the metered data;
17. Rate classification; and
18. The size and operational characteristics of the market in which the program is to operate including the number of covered technologies operating in the market and their expected normal failure, change-out or replacement rates.

Residential program data requests for end-user focused programs

1. Name of program(s) or program component(s) of the participation;
2. Type of building or structure associated with the participant or the participation;
3. Pre-participation measure and measure use information, descriptions and conditions;
4. Service turn on date;
5. Name of individual enrolling in the program or receiving service;
6. Address of the participant;
7. Extent to which customer is a repeat participant or a participant in other programs over the previous five years, if available or accessible;

8. Address where action is taken or for the services received;
9. Listing or description of actions taken or services received according to standard measure and end-use definitions;
10. Individual participation contact information to include:
 - a. First and last name;
 - b. Address;
 - c. Phone number;
 - d. Fax number;(if available and collected); and
 - e. E-mail address (if available and collected).
11. Dates of key action/activity/installation steps associated with program participation:
 - a. Program enrollment date(s);
 - b. Rebate or incentive payment date(s);
 - c. Measure install dates;
 - d. Date of training received; and
 - e. Post-installation inspection dates.
12. Financial assistance amounts paid to participant by measure or action taken;
13. Project description information;
14. Estimated savings for actions taken;
15. Account numbers and meter numbers and consumption histories from utility bills for all relevant meters for at least twelve months prior to program enrollment date and through to current. Note: The evaluation contractor will work with the IOUs to understand what metered data is available for which types of customers and the formats and time intervals associated with the metered data;
16. Rate classification; and
17. The size and operational characteristics of the market in which the program is to operate including the number of covered technologies operating in the market and their expected normal failure, change-out or replacement rates.

Non-participant or rejecter data for end-user focused programs

1. Description of program services offered to customer;
2. Date of offering or contact;
3. Method of contact;
4. Name of contact;
5. Address of contact;
6. Phone number of contact (if known); and
7. E-mail of contact (if known).

Program data for mid-stream and upstream focused programs

1. Name of program(s) or program component(s);
2. Name of firms participating in program or program component;
3. Primary and secondary NAIC codes associated with the participants if available;

4. Extent to which customer is a repeat participant or a participant in other programs over the previous five years, if available or accessible;
5. Pre participation/measure and measure use information, descriptions and conditions;
6. Address of the participating firms or key participation decision makers;
7. Address(es) where action is taken or for the services received;
8. Listing or description of actions taken or services received for each location;
9. Individual participation contact information to include:
 - a. First and last name (if known);
 - b. Address;
 - c. Phone number;
 - d. Fax number (if collected); and
 - e. E-mail address (if collected).
10. Dates of key action/activity/installation steps associated with program participation:
 - a. Program enrollment date(s);
 - b. Rebate or incentive payment date(s);
 - c. Date of training received; and
 - d. Dates, numbers and types of material received.
11. Financial assistance amounts paid to participant by action taken;
12. End-user information as is made available to the program;
13. The size and operational characteristics of the market in which the program is to operate including the number of covered technologies operating in the market and their expected normal failure, change-out or replacement rates; and
14. Names and copies of previous evaluations and market research efforts used by the program to plan and structure program offerings and implementation efforts.

Program data for information, education and advertising-focused programs

1. Name of program(s) or program component(s);
2. Target population description, size, source of identifying information and lists of population members used in outreach activities. The size and operational characteristics of the market in which the program is to operate including the number of covered technologies operating in the market and their expected normal failure, change-out or replacement rates;
3. Contact information where individual participants are identified to include:
 - a. First and last name of key contacts for each location (if known);
 - b. Address of individual contacts;
 - c. Phone number of individual contacts;
 - d. Fax number of individuals (if collected); and
 - e. E-mail address of individuals (if collected).
4. Marketing materials by numbers, types and distribution;
5. Education or Media plan as is appropriate;

6. Execution records for training held; information venues used; program participation agreements, commitments or other similar agreements; post-buy analysis; and other documentation of actual output;
7. Records for dates, number, location, target audience and attendance of events held, Web site hits, call-in numbers and rates, reach, frequency, Gross Rating Points, impressions, click through rate, composition, coverage, earned media, value of public service announcements, and other tracking and monitoring information the program maintains, as appropriate to the effort and for each wave, campaign and targeted effort. Include definitions and calculation methods for monitoring statistics used;
8. End-user information available to the program; and
9. Names and copies of previous evaluations and market research efforts used by the program to plan and structure program offerings and implementation efforts.

Storage and Disposal of Customer Information Used in the Evaluation

Customer information received to support the evaluation efforts will be maintained for a limited period of time consistent with the needs of the evaluation efforts and to support time-series or time-sensitive analysis, but will not be maintained indefinitely. All customer-specific information will be maintained and protected from disclosure for as long as there is an evaluation plan covering the use of the data to support an evaluation effort. Once there is no evaluation plan covering the use of the customer-specific data, it will be deleted or discarded within 3 years in the following ways:

1. Electronic files will be deleted from electronic storage systems;
2. Hard copy files will be shredded and recycled or discarded if it cannot be recycled;
3. Electronic medium (e.g., DVDs, CD, electronic tape) will be shredded and recycled or discarded if it cannot be recycled; and
4. Other materials containing customer-specific information will be rendered unreadable and recycled or discarded.

APPENDIX A. Measure-Level M&V Results Reporting Requirements

Measure-level results from M&V studies shall be reported as unit savings estimates (kWh/unit, kW/unit, therm/unit) normalized in terms consistent with DEER as described below or as amended by CPUC-approved revision of this Appendix.

Table 24. Measure-Level Impact Reporting Requirements

Measure Category	Measure Subcategory	Normalization Units	Measures included
Agricultural	Irrigation	Nozzle	Low pressure nozzles
		Acre of land	Micro irrigation conversion
	Greenhouse	Sq ft of glazing	Heat curtain, IR film
Clothes Dryer	Efficient dryer	Dryer	Efficient clothes dryers
Commercial Cooking	Equipment	Equipment	Griddles, fryers, warming cabinets, steamers
Commercial Refrigeration	Controls	Design evaporator tons	Floating head pressure and suction pressure controls
		Case linear feet	Case lighting controls
		Motor	Case and cooler fan controls
	Equipment	1000 SF of sales area	Refrigerant holdback valves
		Design compressor tons	VSD on compressor
		Design evaporator tons	Oversized condensers, sub cooling, compressor and condenser change outs
		Case linear feet	Case covers, reach-in conversions, case replacements
		Motor	Efficient evaporator fan motors
		Cooler	Door closers
		Door	Anti-sweat heater elimination
		Freezer	Door closers
	Maintenance	Design evaporator tons	Refrigeration system maintenance
Hot Water	Circulation Pump	1000 SF of building	Hot water recirculation pumps and pump controls
	Clothes washer	Clothes washer	Efficient clothes washers
	Dishwasher	Dishwasher	Efficient dishwashers
	Faucet aerators	Household	Faucet aerators
	Heat pump water heater	Water heater	Heat pump water heaters
	Efficient water heater - residential	Water heater	Storage type water heaters
	Efficient water heater - commercial	1000 SF of building	Storage type water heaters
	Low flow showerhead	Showerhead	Low flow showerheads

Evaluators' Protocols

Measure Category	Measure Subcategory	Normalization Units	Measures included
	Pipe wrap	Household	Hot water pipe insulation and heat traps
	Point of use water heater	1000 SF of building	Instantaneous water heaters
HVAC	Controls	Tons	Economizer retrofit and repair
		1000 SF of building	Programmable thermostats, energy management systems, time clocks, space heating hot water and chilled water loop temperature control
	Equipment	1000 SF of building	Ventilation rate changes, evaporative coolers, air to air heat recovery, commercial furnaces and boilers
		kBtu/hr of furnace capacity	Residential furnaces
		Nameplate motor hp	Efficient motors in HVAC applications, 3 way to 2 way valve conversions on chilled water and space heating hot water coils, variable frequency drives
Lighting		Tons	High efficiency packaged AC and heat pumps, high efficiency chillers, waterside economizers, evaporative ventilation air pre-coolers
	Ballast	Fixture	Dimming ballasts
	CFL lamps	Lamp	Screw-in and hardwire compact fluorescent lamps
	De-lamp	Fixture	All interior lighting fixture types
	Exit sign	Exit sign	LED exit signs
	Exterior lighting	Lamp	HID lamps for exterior lighting applications
	Linear fluorescent	Fixture	High efficiency fluorescent lighting fixtures with T-8 or T-5, linear or U-Tube lamps
	Metal halide	Lamp	High efficiency metal halide lamps
	Occupancy sensor	Sensor	Occupancy sensors for interior lighting applications
	Photocell	Photocell	Photocell controls for exterior lighting applications
	Time clock	Time clock	Time clock controls for interior or exterior lighting applications
	Lighting controls - general	kW controlled	Other general purpose interior lighting control systems
	LPD reduction	kW reduced	Efficient lighting design providing reduced lighting power density
Interior Plug Loads	Copy machine	Copy machine	Efficient copy machine
	Equipment	kW reduced	Use of efficient office equipment resulting in equipment power density reduction
Miscellaneous	Motors	Motor	Efficient non-HVAC motors
	Vending machine	Machine	Efficient vending machines and vending machine controllers
Pools	Pool pump	Pump	Efficient pool pumps and pool pump controllers

Evaluators' Protocols

Measure Category	Measure Subcategory	Normalization Units	Measures included
Residential Refrigeration	Refrigerator	Refrigerator	Efficient residential refrigerators or refrigerator/freezers
	Freezer	Freezer	Efficient residential freezers
Shell	Shell	1000 SF of building	Weatherization, air leakage sealing
	Fenestration	100 SF of window	High performance windows, skylights and glazing systems
	Insulation	1000 SF of insulation	Insulation, cool roofs
	Equipment	1000 SF of building	Whole-house fans

APPENDIX B. Glossary of Terms¹⁴¹

ACCURACY - An indication of how close a value is to the true value of the quantity in question. The term could also be used in reference to a model or a set of measured data, or to describe a measuring instrument's capability.

ACHIEVABLE POTENTIAL - The amount of savings that can be achieved due to specific program designs and delivery approaches, including program funding and measure incentive levels. Achievable potential studies are sometimes referred to as Market Potential studies.

ADMINISTRATOR - A person, company, partnership, corporation, association or other entity selected by the CPUC and any subcontractor that is retained by an aforesaid entity to contract for and administer energy efficiency programs funded in whole or in part from electric or gas Public Goods Charge (PGC) funds. For purposes of implementing PU Code Section 381.1, an "administrator" is any party that receives funding for and implements energy efficiency programs pursuant to PU Code Section 381. Similarly, a person, company or other entity selected to contract and administer energy efficiency programs funded by procurement funds.

ANALYSIS OF COVARIANCE (ANCOVA) MODELS - A type of regression model also referred to as a "fixed effects" model. This model allows each individual to act as its own control. The unique effects of the stable, but unmeasured characteristics of each customer are their "fixed effects" from which this method takes its name. These fixed effects are held constant.

ASHRAE - American Society of Heating, Refrigerating and Air- Conditioning Engineers

AUTOCORRELATION - The breakdown in the assumptions that the errors in regression analysis are uncorrelated due to correlation in the error term across observations in a time-series or cross-series, the error in one time period is directly correlated to the error in another time period or cross-sectional category. First-order serial correlation is when that correlation is with the error in the subsequent/preceding time period. The correlation can be positive or negative.

BASELINE DATA - The measurements and facts describing facility operations and design during the baseline period. This will include energy use or demand and parameters of facility operation that govern energy use or demand.

BASELINE FORECAST - A prediction of future energy needs which does not take into account the likely effects of new efficiency programs that have not yet been started.

BASELINE MODEL - The set of arithmetic factors, equations or data used to describe the relationship between energy use or demand and other baseline data. A model may also be a simulation process involving a specified simulation engine and set of input data.

¹⁴¹ Terms defined as used herein and within the context of energy efficiency evaluation.

BASELINE PERIOD - The period of time selected as representative of facility operations before retrofit.

BEHAVIORAL DEGRADATION FACTOR - A multiplier used to account for time-related change in the energy savings of a high efficiency measure or practice relative to a standard efficiency measure or practice due to changes in behavior in relation to the measure or practice.

BILLING DATA - Has multiple meanings. Metered data obtained from the electric or gas meter used to bill the customer for energy used in a particular billing period. Meters used for this purpose typically conform to regulatory standards established for each customer class. Also used to describe the data representing the bills customers receive from the energy provider and also used to describe the customer billing and payment streams associated with customer accounts. This term is used to describe both consumption and demand, and account billing and payment information.

BILLING DEMAND - The demand used to calculate the demand charge cost. This is very often the monthly peak demand of the customer, but it may have a floor of some percentage of the highest monthly peak of the previous several months (a demand “ratchet”). May have other meanings associated with customer account billing practices.

BRITISH THERMAL UNIT (Btu or BTU) - The standard measure of heat energy. It takes one Btu to raise the temperature of one pound of water by one degree Fahrenheit at sea level. For example, it takes about 1,000 BTUs to make a pot of coffee. One Btu is equivalent to 252 calories, 778 foot-pounds, 1055 joules and 0.293 watt-hours. Note: the abbreviation is seen as “Btu” or “BTU” interchangeably.

BUILDING COMMISSIONING - Building commissioning provides documented confirmation that building systems as constructed function in accordance with the intent of the building designers and satisfy the owner’s operational needs.

BUILDING ENERGY EFFICIENCY STANDARDS - California Code of Regulations, Title 24, Part 2, Chapter 2-53; regulating the energy efficiency of buildings constructed in California.

BUILDING ENERGY SIMULATION MODEL - Computer models based on physical engineering principals and/or standards used to estimate energy usage and/or savings. These models do not make use of billing or metered data, but usually incorporate site-specific data on customers and physical systems. Building Simulation Models usually require such site-specific data as square footage, weather, surface orientations, elevations, space volumes, construction materials, equipment use, lighting and building occupancy. Building simulation models can usually account for interactive effects between end-uses (e.g., lighting and HVAC), part-load efficiencies and changes in external and internal heat gains/losses. Examples of building simulation models include ADM2, BLAST and DOE-2.

BUILDING ENVELOPE - The assembly of exterior partitions of a building that enclose conditioned spaces, through which thermal energy may be transferred to or from the exterior,

unconditioned spaces or the ground. (See California Code of Regulations, Title 24, Section 2-5302.)

CADMAC - See CALIFORNIA DEMAND-SIDE MANAGEMENT MEASUREMENT ADVISORY COUNCIL.

CALIFORNIA MEASUREMENT ADVISORY COUNCIL (CALMAC) - An informal committee made up of representatives of the California IOUs, CPUC, CEC and NRDC. CALMAC provides a forum for the development, implementation, presentation, discussion and review of regional and statewide market assessment and evaluation studies for California energy efficiency programs conducted by member organizations using Public Goods Charge funds.

CALIFORNIA DEMAND-SIDE MANAGEMENT MEASUREMENT ADVISORY COUNCIL (CADMAC) - An informal committee made up of utility representatives, the Office of Ratepayer Advocates and the CEC. The purpose of the committee is to: provide a forum for presentations, discussions and review of Demand Side Management (DSM) program measurement studies underway or completed; to coordinate the development and implementation of measurement studies common to all or most of the utilities; and to facilitate the development of effective, state-of-the-art Protocols for measuring and evaluating the impacts of DSM programs.

CALIFORNIA ENERGY COMMISSION (CEC) - The state agency established by the Warren-Alquist State Energy Resources Conservation and Development Act in 1974 (Public Resources Code, Sections 25000 et seq.) responsible for energy policy. Funding for the CEC's activities comes from the Energy Resources Program Account, Federal Petroleum Violation Escrow Account and other sources. The CEC has statewide power plant siting, supply and demand forecasting, as well as multiple types of energy policy and analysis responsibilities.

CALIFORNIA PUBLIC UTILITIES COMMISSION (CPUC) - A state agency created by constitutional amendment in 1911 to regulate the rates and services of privately owned utilities and transportation companies. The CPUC is an administrative agency that exercises both legislative and judicial powers; its decisions and orders may be appealed only to the California Supreme Court. The major duties of the CPUC are to regulate privately owned utilities, securing adequate service to the public at rates that are just and reasonable both to customers and shareholders of the utilities; including rates, electricity transmission lines and natural gas pipelines. The CPUC also provides electricity and natural gas forecasting, and analysis and planning of energy supply and resources. Its headquarters are in San Francisco.

CALMAC – See CALIFORNIA MANAGEMENT MEASUREMENT ADVISORY COUNCIL.

CAPACITY - The amount of electric power for which a generating unit, generating station or other electrical apparatus is rated either by the user or manufacturer. The term is also used for the total volume of natural gas that can flow through a pipeline over a given amount of time, considering such factors as compression and pipeline size.

CEC - See CALIFORNIA ENERGY COMMISSION.

CHANGE MODEL - A type of billing analysis designed to explain changes in energy usage. This can take the form of having the change in energy consumption (pre versus post) as the dependent variable (e.g., December pre-retrofit usage – December post-retrofit usage) or having consumption as the dependent variable and pre-retrofit consumption as one of the independent variables.

CLIMATE THERMAL ZONE (CTZ) – A geographical area in the state that has particular weather patterns. These zones are used to determine the type of building standards that are required by law.

CLTD – See COOLING LOAD TEMPERATURE DIFFERENCE.

COEFFICIENT OF VARIATION - The sample standard deviation divided by the sample mean ($cv = sd/y$). See page 320 of the *Evaluation Framework*.

COINCIDENT DEMAND - The metered demand of a device, circuit or building that occurs at the same time as the peak demand of the building or facility or at the same time as some other peak of interest, such as a utility's system load. This should properly be expressed so as to indicate the peak of interest, e.g., "demand coincident with the building peak."

COMMERCIALIZATION - Programs or activities that increase the value or decrease the cost of integrating new products or services into the electricity sector.

COMMISSIONING - See BUILDING COMMISSIONING.

COMPARISON GROUP - A group of customers who did not participate in a given program during the program year and who share as many characteristics as possible with the participant group.

COMPREHENSIVE - A program or project designed to achieve all cost-effective energy efficiency activities in individual buildings, usually including multiple energy efficiency measures.

CONDITIONAL DEMAND ANALYSIS (CDA) - A type of billing analysis in which observed energy consumption is estimated as a function of major end-uses, often portrayed as dummy variables for their existence at the customer residence/facility.

CONDITIONED FLOOR AREA - The floor area of enclosed conditioned spaces on all floors measured from the interior surfaces of exterior partitions for nonresidential buildings and from the exterior surfaces of exterior partitions for residential buildings. (See California Code of Regulations, Title 24, Section 2-5302.)

CONDITIONED SPACE - Enclosed space that is either directly or indirectly conditioned. (See California Code of Regulations, Title 24, Section 2-5302.)

CONDITIONED SPACE, DIRECTLY - An enclosed space that is provided with heating equipment that has a capacity exceeding 10 Btus/(hr-ft²) or with cooling equipment that has a capacity exceeding 10 Btus/(hr-ft²). An exception is if the heating and cooling equipment is designed and thermostatically controlled to maintain a process environment temperature less than 65° F or greater than 85° F for the whole space the equipment serves. (See California Code of Regulations, Title 24, Section 2- 5302.)

CONDITIONED SPACE, INDIRECTLY - Enclosed space that: (1) has a greater area weighted heat transfer coefficient (u-value) between it and directly conditioned spaces than between it and the outdoors or unconditioned space; (2) has air transferred from directly conditioned space moving through it at a rate exceeding three air changes/hour.

CONSERVATION - Steps taken to cause less energy to be used than would otherwise be the case. These steps may involve, for example, improved efficiency, avoidance of waste, and reduced consumption. Related activities include, for example, installing equipment (such as a computer to ensure efficient energy use), modifying equipment (such as making a boiler more efficient), adding insulation and changing behavior patterns.

CONSTRUCT VALIDITY - The extent to which an operating variable/instrument accurately taps an underlying concept/hypothesis, properly measuring an abstract quality or idea.

CONTENT VALIDITY - The extent to which an operating measure taps all the separate sub-concepts of a complicated concept.

CONVERGENT VALIDITY - When two instruments/questions/measurement methods obtain similar results when measuring the same underlying construct with varying questions/approaches.

COOLING DEGREE DAYS - The cumulative number of degrees in a month or year by which the mean temperature is above 18.3°C /65° F.

CORRELATION COEFFICIENT - A measure of the linear association between two variables, calculated as the square root of the R² obtained by regressing one variable on the other and signed to indicate whether the relationship is positive or negative.

COST-EFFECTIVENESS - An indicator of the relative performance or economic attractiveness of any energy efficiency investment or practice when compared to the costs of energy produced and delivered in the absence of such an investment. In the energy efficiency field, the present value of the estimated benefits produced by an energy efficiency program as compared to the estimated total program's costs, from the perspective of either society as a whole or of individual customers, to determine if the proposed investment or measure is desirable from a variety of perspectives, e.g., whether the estimated benefits exceed the estimated costs. See also TOTAL RESOURCE COST TEST – SOCIETAL VERSION and PARTICIPANT COST TEST.

CPUC - See CALIFORNIA PUBLIC UTILITIES COMMISSION.

CTZ – See CLIMATE THERMAL ZONE.

CUSTOMER - Any person or entity responsible for payment of an electric and/or gas bill to and with an active meter serviced by a utility company (refers to IOU customers herein).

CUSTOMER INFORMATION - Non-public information and data specific to a utility customer that the utility acquired or developed in the course of its provision of utility services.

CV – See COEFFICIENT OF VARIATION.

DATABASE FOR ENERGY-EFFICIENT RESOURCES (DEER) – A database sponsored by the CEC and CPUC designed to provide well-documented estimates of energy and peak demand savings values, measure costs, and effective useful life (EUL) all with one data source. The users of the data are intended to be program planners, regulatory reviewers and planners, utility and regulatory forecasters, and consultants supporting utility and regulatory research and evaluation efforts. DEER has been designated by the CPUC as its source for deemed and impact costs for program planning.

DAYLIGHTING - The use of sunlight to supplement or replace electric lighting.

DEER – See DATABASE FOR ENERGY-EFFICIENT RESOURCES.

DEMAND - The time rate of energy flow. Demand usually refers to electric power and is measured in kW (equals kWh/h) but can also refer to natural gas, usually as Btu/hr, kBtu/hr, therms/day or ccf/day.

DEMAND (Utility) - The rate or level at which electricity or natural gas is delivered to users at a given point in time. Electric demand is expressed in kilowatts (kW). Demand should not be confused with load, which is the amount of power delivered or required at any specified point or points on a system.

DEMAND BILLING - The electric capacity requirement for which a large user pays. It may be based on the customer's peak demand during the contract year, on a previous maximum or on an agreed minimum. Measured in kilowatts.

DEMAND CHARGE - The sum to be paid by a large electricity consumer for its peak usage level.

DEMAND RESPONSIVENESS - Also sometimes referred to as load shifting. Activities or equipment that induce consumers to use energy at different (lower cost) times of day or to interrupt energy use for certain equipment temporarily, usually in direct response to a price signal. Examples include interruptible rates, doing laundry after 7 p.m., and air conditioner recycling programs.

DEMAND SAVINGS - The reduction in the demand from the pre-retrofit baseline to the post-retrofit demand, once independent variables (such as weather or occupancy) have been adjusted

for. This term is usually applied to billing demand, to calculate cost savings or to peak demand, for equipment sizing purposes.

DEMAND SIDE MANAGEMENT (DSM) - The methods used to manage energy demand including energy efficiency, load management, fuel substitution and load building. See **LOAD MANAGEMENT**.

DIRECT ENERGY SAVINGS (DIRECT PROGRAM ENERGY SAVINGS) - The use of the words “direct savings” or “direct program savings” refers to the savings from programs that are responsible for the achievement of specific energy efficiency goals. Typically these are thought of as resource acquisition programs or programs that install or expedite the installation of energy-efficient equipment and which directly cause or help to cause energy efficiency to be achieved. Rebate, incentive or direct install programs provide direct energy savings.

DIRECT INSTALL or DIRECT INSTALLATION PROGRAMS - These types of programs provide free energy efficiency measures and their installation for qualified customers. Typical measures distributed by these programs include low flow showerheads and compact fluorescent bulbs.

DIRECTLY COOLED SPACE is an enclosed space that is provided with a space-cooling system that has a capacity exceeding 5 Btu/(hr×ft²), unless the space-cooling system is designed and thermostatically controlled to maintain a space temperature less than 55°F or to maintain a space temperature greater than 90°F for the whole space that the system serves.

DIRECTLY HEATED SPACE is an enclosed space that is provided with wood heating or is provided with a space-heating system that has a capacity exceeding 10 Btu/(hr×ft²) unless the space-heating system is designed and thermostatically controlled to maintain a space temperature less than 55°F or to maintain a space temperature greater than 90°F for the whole space that the system serves.

DISTRIBUTED GENERATION - A distributed generation system involves small amounts of generation located on a utility's distribution system for the purpose of meeting local (substation level) peak loads and/or displacing the need to build additional (or upgrade) local distribution lines.

DOUBLE-BARRELED QUESTIONS - A poorly worded questionnaire item, which actually asks two questions at the same time, thereby not allowing unique and accurate interpretation of the results.

DRY-BULB TEMPERATURE - A measure of the sensible temperature of air.

DSM - See **DEMAND SIDE MANAGEMENT**.

ECM – Energy Conservation Measure. See **MEASURE** and **ENERGY EFFICIENCY MEASURE**.

EDUCATION PROGRAMS - Programs primarily intended to educate customers about energy-efficient technologies or behaviors or provide information about programs that offer energy efficiency or load reduction information or services.

EFFECTIVE USEFUL LIFE (EUL) - An estimate of the median number of years that the measures installed under a program are still in place and operable.

EFFICIENCY - The ratio of the useful energy delivered by a dynamic system (such as a machine, engine or motor) to the energy supplied to it over the same period or cycle of operation. The ratio is usually determined under specific test conditions.

ELECTRIC PUBLIC GOODS CHARGE (PGC) - Per Assembly Bill (AB) 1890, a universal charge applied to each electric utility customer's bill to support the provision of public goods. Public goods covered by California's electric PGC include public purpose energy efficiency programs, low-income services, renewables, and energy-related research and development.

EM&V - Evaluation, Measurement, Monitoring and Verification.

EMISSIVITY - The property of emitting radiation; possessed by all materials to a varying extent.

EMITTANCE - The emissivity of a material, expressed as a fraction. Emittance values range from 0.05 for brightly polished metals to 0.96 for flat black paint.

END-USE (MEASURES/GROUPS) - Refers to a broad or sometimes narrower category that the program is concentrating efforts upon. Examples of end-uses include refrigeration, food service, HVAC, appliances, envelope and lighting.

ENERGY CONSUMPTION - The amount of energy consumed in the form in which it is acquired by the user. The term excludes electrical generation and distribution losses.

ENERGY COST - The total cost for energy, including such charges as base charges, demand charges, customer charges, power factor charges and miscellaneous charges.

ENERGY EFFICIENCY - Using less energy to perform the same function. Programs designed to use energy more efficiently - doing the same with less. For the purpose of this paper, energy efficiency programs are distinguished from DSM programs in that the latter are utility-sponsored and financed, while the former is a broader term not limited to any particular sponsor or funding source. "Energy conservation" is a term that has also been used but it has the connotation of doing without in order to save energy rather than using less energy to perform the same function and so is not used as much today. Many people use these terms interchangeably.

ENERGY EFFICIENCY IMPROVEMENT - Reduced energy use for a comparable level of service, resulting from the installation of an energy efficiency measure or the adoption of an energy efficiency practice. Level of service may be expressed in such ways as the volume of a

refrigerator, temperature levels, production output of a manufacturing facility or lighting level/square foot.

ENERGY EFFICIENCY MEASURE - Installation of equipment, subsystems or systems, or modification of equipment, subsystems, systems or operations on the customer side of the meter, for the purpose of reducing energy and/or demand (and, hence, energy and/or demand costs) at a comparable level of service.

ENERGY EFFICIENCY OF A MEASURE - A measure of the energy used to provide a specific service or to accomplish a specific amount of work (e.g., kWh/cubic foot of a refrigerator, therms/gallon of hot water).

ENERGY EFFICIENCY OF EQUIPMENT - The percentage of gross energy input that is realized as useful energy output of a piece of equipment.

ENERGY EFFICIENCY PRACTICE - The use of high-efficiency products, services and practices or an energy-using appliance or piece of equipment, to reduce energy usage while maintaining a comparable level of service when installed or applied on the customer side of the meter. Energy efficiency activities typically require permanent replacement of energy-using equipment with more efficient models. Examples: refrigerator replacement, light fixture replacement, cooling equipment upgrades.

ENERGY MANAGEMENT SYSTEM - A control system (often computerized) designed to regulate the energy consumption of a building by controlling the operation of energy consuming systems, such as the heating, ventilation and air conditioning (HVAC), lighting and water heating systems.

ENERGY RESOURCES PROGRAM ACCOUNT (ERPA) - The state law that directs California electric utility companies to gather a state energy surcharge/kilowatt-hour of electricity consumed by a customer. These funds are used for operation of the CEC. As of January 1, 2004, the surcharge is set at \$0.0003/kWh.

ENERGY SAVINGS - The reduction in use of energy from the pre-retrofit baseline to the post-retrofit energy use, once independent variables (such as weather or occupancy) have been adjusted for.

ENGINEERING APPROACHES - Methods using engineering algorithms or models to estimate energy and/or demand use.

ENGINEERING USEFUL LIFE - An engineering estimate of the number of years that a piece of equipment will operate if properly maintained.

ERPA - See **ENERGY RESOURCES PROGRAM ACCOUNT**.

ERROR - Deviation of measurements from the true value.

EUL - See EFFECTIVE USEFUL LIFE.

EVALUATION - The performance of studies and activities aimed at determining the effects of a program; any of a wide range of assessment activities associated with understanding or documenting program performance or potential performance, assessing program or program-related markets and market operations; any of a wide range of evaluative efforts including assessing program-induced changes in energy efficiency markets, levels of demand or energy savings and program cost-effectiveness.

EX-ANTE SAVINGS ESTIMATE – Administrator-forecasted savings used for program and portfolio planning purposes as filed with the CPUC, from the Latin for “beforehand.”

EX-POST EVALUATION ESTIMATED SAVINGS - Savings estimates reported by the independent evaluator after the energy impact evaluation and the associated M&V efforts have been completed. If only the term “ex-post savings” is used, it will be assumed that it is referring to the ex-post evaluation estimate, the most common usage, from the Latin for “from something done afterward.”

EX-POST (PROGRAM) ADMINISTRATOR-ESTIMATED SAVINGS - Savings estimates reported by the Administrator after program implementation has begun (Administrator-reported ex post), from the Latin for “from something done afterward.”

EX-POST (PROGRAM) ADMINISTRATOR-FORECASTED SAVINGS – Savings estimates forecasted by the Administrator during the program and portfolio planning process, from the Latin for “from something done afterward.”

EXTERNAL VALIDITY - The extent to which the association between an independent variable and a dependent variable that is demonstrated within a research setting also holds true in the general environment.

FREE-DRIVER - A non-participant who adopted a particular efficiency measure or practice as a result of a utility program. See SPILLOVER EFFECTS for aggregate impacts.

FREE-RIDER - A program participant who would have implemented the program measure or practice in the absence of the program.

GAS PUBLIC GOODS CHARGE (PGC) - Created by AB1002 in 2000, an unbundled rate component included on gas customer bills to fund public purpose programs including those for energy efficiency, low-income, and research and development.

GIGAWATT (GW) - One thousand megawatts (1,000 MW), one million kilowatts (1,000,000 kW) or one billion watts (1,000,000,000 watts) of electricity. One gigawatt is enough to supply the electric demand of about one million average California homes.

GIGAWATT-HOUR (GWH) - One million kilowatt-hours of electric power.

GLAZING - A covering of transparent or translucent material (typically glass or plastic) used for admitting light.

GROSS AREA - The area of a surface including areas not belonging to that surface (such as windows and doors in a wall).

GROSS LOAD IMPACT - The change in energy consumption and/or demand that results directly from program-related actions taken by participants in a DSM program, regardless of why they participated. Related to Gross Energy Impact and Gross Demand Protocols.

HEAT CAPACITY - The amount of heat necessary to raise the temperature of a given mass one degree. Heat capacity may be calculated by multiplying the mass by the specific heat.

HEAT GAIN - An increase in the amount of heat contained in a space, resulting from direct solar radiation, heat flow through walls, windows and other building surfaces, and the heat given off by people, lights, equipment and other sources.

HEAT LOSS - A decrease in the amount of heat contained in a space, resulting from heat flow through walls, windows, roof and other building surfaces and from exfiltration of warm air.

HEAT PUMP - An air conditioning unit which is capable of heating by refrigeration, transferring heat from one (often cooler) medium to another (often warmer) medium and which may or may not include a capability for cooling. This reverse-cycle air conditioner usually provides cooling in summer and heating in winter.

HEAT TRANSFER - Flow of heat energy induced by a temperature difference. Heat flow through a building envelope typically flows from a heated or hot area, to a cooled or cold area.

HETEROSCEDASTICITY – Unequal error variance. In statistics, a sequence or a vector of random variables is heteroscedastic if the random variables in the sequence or vector may have different variances. This violates the regression assumption of constant variance (the variance of the errors is constant across observations or homoscedastic). Typically, residuals are plotted to assess this assumption. Standard estimation methods are inefficient when the errors are heteroscedastic. A common example is when variance is expected to be greater on a variable measurement for larger firms than for smaller firms.

HOMOSCEDASTIC (HOMOSCEDASTICITY) - Constant error variance, an assumption of classical regression analysis. See also HETEROSCEDASTICITY.

HORSEPOWER (HP) - A unit for measuring the rate of doing work. One horsepower equals about three-fourths of a kilowatt (745.7 watts).

HVAC - Heating Ventilation and Air Conditioning.

HVAC SYSTEM - The equipment, distribution network and terminals that provides either collectively or individually the processes of heating, ventilating or air conditioning to a building.

IMPACT EVALUATION - Used to measure the program-specific induced changes in energy and/or demand usage (such kWh, kW and therms) and/or behavior attributed to energy efficiency and demand response programs.

IMPACT YEAR - Depending on the context, impact year means either (a) the twelve months subsequent to program participation used to represent program costs or load impacts occurring in that year, or (b) any calendar year after the program year in which impacts may occur.

IMPLEMENTATION THEORY - A theory describing how a program should be structured and implemented and the theoretical rationale supporting the reasons for the program structure and the implementation approach.

IMPLEMENTER - An entity or person selected and contracted with or qualified by a program Administrator or by the CPUC to receive PGC funds for providing products and services to customers.

INCENTIVES - Financial support (e.g., rebates, low-interest loans) to install energy efficiency measures. The incentives are solicited by the customer and based on the customer's billing history and/or customer-specific information.

INDEPENDENT VARIABLES - The factors that affect the energy and demand used in a building but cannot be controlled (e.g., weather or occupancy).

INDIRECT ENERGY SAVINGS (INDIRECT PROGRAM ENERGY SAVINGS) - The use of the words "indirect savings" or "indirect program savings" refers to programs that are typically information, education, marketing or outreach programs in which the program's actions are expected to result in energy savings achieved through the actions of the customers exposed to the program's efforts, without direct enrollment in an program that has energy savings goals.

INFORMATION PROGRAMS - Programs primarily intended to provide customers with information regarding generic (not customer-specific) conservation and energy efficiency opportunities. For these programs, the information may be unsolicited by the customer. Programs that provide incentives in the form of unsolicited coupons for discount on low cost measures are also included.

INSULATION, THERMAL - A material having a relatively high resistance of heat flow and used principally to retard heat flow. See R-VALUE.

INTEGRATED PART-LOAD VALUE (IPLV) - A single number figure of merit based on part-load EER or COP expressing part-load efficiency for air conditioning and heat pump equipment on the basis of weighted operation at various load capacities for the equipment.

INTERNAL VALIDITY - The extent to which alternative explanations can be eliminated as causes for an observed association between independent and dependent variable(s) within a research setting/sample.

INTERNATIONAL PERFORMANCE MEASUREMENT AND VERIFICATION PROTOCOL (IPMVP) – The IPMVP provides an overview of current best practice techniques available for verifying results of energy efficiency, water efficiency, and renewable energy projects in commercial and industrial facilities. It may also be used by facility operators to assess and improve facility performance. The IPMVP is the leading international standard in M&V protocols. It has been translated into 10 languages and is used in more than 40 countries.

INVESTOR-OWNED UTILITY (IOU) - A private company that provides a utility, such as water, natural gas or electricity, to a specific service area. California investor-owned utilities are regulated by the CPUC.

IPMVP – See INTERNATIONAL PERFORMANCE MEASUREMENT AND VERIFICATION PROTOCOL.

JOULE - A unit of work or energy equal to the amount of work done when the point of application of force of 1 Newton is displaced 1 meter in the direction of the force. It takes 1,055 joules to equal a British thermal unit. It takes about 1 million joules to make a pot of coffee.

kBtu – One thousand (1,000) British Thermal Units (Btu). See also BRITISH THERMAL UNIT.

KILOWATT (kW) - One thousand (1,000) watts. A unit of measure of the amount of electricity needed to operate given equipment. On a hot summer afternoon a typical home with central air conditioning and other equipment in use might have a demand of four kW each hour.

KILOWATT-HOUR (kWh) - The most commonly used unit of measure indicating the amount of electricity consumed over time; one kilowatt of electricity supplied for one hour.

LEVEL OF SERVICES - The utility received by a customer from energy-using equipment. Level of service may be expressed, for example, as the volume of a refrigerator, an indoor temperature level, the production output of a manufacturing facility, or lighting levels/square foot.

LOAD - The amount of electric power supplied to meet one or more end-user's needs. The amount of electric power delivered or required at any specified point or points on a system. Load originates primarily at the power-consuming equipment of the customer. Load should not be confused with demand, which is the rate at which power is delivered to or by a system, part of a system, or a piece of equipment.

LOAD DIVERSITY - The condition that exists when the peak demands of a variety of electric customers occur at different times. The difference between the peak of coincident and noncoincident demands of two or more individual loads. This is the objective of "load molding" strategies, ultimately curbing the total capacity requirements of a utility.

LOAD FACTOR - The ratio of the amount of electricity a consumer used during a given time span and the amount that would have been used if the usage had stayed at the consumer's highest

demand level during the whole time. The term also is used to mean the percentage of capacity of an energy facility - such as a power plant or gas pipeline - that is utilized in a given period of time. The ratio of the average load to peak load during a specified time interval.

LOAD IMPACT - Changes in electric energy use, electric peak demand or natural gas use.

LOAD MANAGEMENT - Steps taken to reduce power demand at peak load times or to shift some power demand to off-peak times to better meet the utility system capability for a given hour, day, week, season, or year. Load management may be obtained by persuading consumers to modify behavior, by using equipment that regulates or controls electric consumption or by other means.

LOAD SHAPE - The time-of-use pattern of customer or equipment energy use. This pattern can be over a day (24 hours) or over a year (8760 hours).

LOAD SHAPE IMPACTS - Changes in load shape induced by a program.

LOGIC MODEL - The graphical representation of the program theory showing the flow between activities, their outputs and subsequent short-term, intermediate and long-term outcomes. Often the logic model is displayed with these elements in boxes and the causal flow being shown by arrows from one to the others in the program logic. It can also be displayed as a table with the linear relationship presented by the rows in the table.

LOW-E - A special coating that reduces the emissivity of a window assembly, thereby reducing the heat transfer through the assembly.

LUMEN - A measure of the amount of light available from a light source equivalent to the light emitted by one candle.

LUMENS/WATT - A measure of the efficacy of a light fixture; the number of lumens output/watt of power consumed.

MAIN METER - The meter that measures the energy used for the whole facility. There is at least one meter for each energy source and possibly more than one per source for large facilities. Typically, utility meters are used, but dataloggers may also be used as long as they isolate the load for the facility being studied. When more than one meter per energy source exists for a facility, the main meter may be considered the accumulation of all the meters involved.

MARKET - The commercial activity (manufacturing, distributing, buying and selling) associated with products and services that affect energy usage.

MARKET ACTORS - Individuals and organizations in the production, distribution and/or delivery chain of energy efficiency products, services and practices. This may include, but is not limited to, manufacturers, distributors, wholesalers, retailers, vendors, dealers, contractors, developers, builders, financial institutions, and real estate brokers and agents.

MARKET ASSESSMENT - An analysis function that provides an assessment of how and how well a specific market or market segment is functioning with respect to the definition of well-functioning markets or with respect to other specific policy objectives. Generally includes a characterization or description of the specific market or market segments, including a description of the types and number of buyers and sellers in the market, the key actors that influence the market, the type and number of transactions that occur on an annual basis and the extent to which energy efficiency is considered an important part of these transactions by market participants. This analysis may also include an assessment of whether or not a market has been sufficiently transformed to justify a reduction or elimination of specific program interventions. Market assessment can be blended with strategic planning analysis to produce recommended program designs or budgets. One particular kind of market assessment effort is a baseline study, or the characterization of a market before the commencement of a specific intervention in the market, for the purpose of guiding the intervention and/or assessing its effectiveness later.

MARKET BARRIER - Any characteristic of the market for an energy-related product, service or practice that helps to explain the gap between the actual level of investment in, or practice of, energy efficiency and an increased level that would appear to be cost-beneficial to the consumer.

MARKET EFFECT - A change in the structure or functioning of a market or the behavior of participants in a market that result from one or more program efforts. Typically these efforts are designed to increase in the adoption of energy-efficient products, services or practices and are causally related to market interventions.

MARKET EVENT - The broader circumstances under which a customer considers adopting an energy efficiency product, service or practice. Types of market events include, but are not necessarily limited to: (a) new construction, or the construction of a new building or facility; (b) renovation, or the updating of an existing building or facility; (c) remodeling, or a change in an existing building; (d) replacement, or the replacement of equipment, either as a result of an emergency such as equipment failure or as part of a broader planned event; and, (e) retrofit, or the early replacement of equipment or refitting of a building or facility while equipment is still functioning, often as a result of an intervention into energy efficiency markets.

MARKET PARTICIPANTS - The individuals and organizations participating in transactions with one another within an energy efficiency market or markets, including customers and market actors.

MARKET POTENTIAL - See **ACHIEVABLE POTENTIAL**.

MARKET SECTORS - General types of markets that a program may target or in which a service offering may be placed. Market sectors include categories such as Agricultural, Commercial, Industrial, Government and Institutional. Market sectors help the CPUC assess how well its portfolio of programs is addressing the variety of markets for energy efficiency products and services in the state.

MARKET SEGMENTS - A part of a market sector that can be grouped together as a result of a characteristic similar to the group. Within the residential sector are market segments such as

renters, owners, multi-family and single-family. These market segments help the CPUC assess how well its portfolio of programs is addressing the variety of segments within the markets served.

MARKET THEORY - A theoretical description of how a market operates relative to a specific program or set of programs designed to influence that market. Market theories typically include the identification of key market actors, information flows and product flows through the market, relative to a program designed to change the way the market operates. Market theories are typically grounded upon the information provided from a market assessment but can also be based on other information. Market theories often describe how a program intervention can take advantage of the structure and function of a market to transform the market. Market theories can also describe the key barriers and benefits associated with a market and describe how a program can exploit the benefits and overcome the barriers.

MARKET TRANSFORMATION - A reduction in market barriers resulting from a market intervention, as evidenced by a set of market effects, that lasts after the intervention has been withdrawn, reduced or changed.

MEASURE (noun) - A product whose installation and operation at a customer's premises results in a reduction in the customer's on-site energy use, compared to what would have happened otherwise. See also **ENERGY EFFICIENCY MEASURE**.

MEASURE (verb) - Use of an instrument to assess a physical quantity or use of a computer simulation to estimate a physical quantity.

MEASURE RETENTION – The degree to which measures are retained in use after they are installed. Measure retention studies assess the length of time the measure(s) installed during the program year are maintained in operating condition and the extent to which there has been a significant reduction in the effectiveness of the measure(s).

MEASURED SAVINGS - Savings or reductions in billing determinants, which are determined using engineering analysis in combination with measured data or through billing analysis.

MEGAWATT (MW) - One thousand kilowatts (1,000 kW) or one million (1,000,000) watts. One megawatt is enough energy to power 1,000 average California homes.

MEGAWATT HOUR (MWh) - One thousand kilowatt-hours. This amount of electricity would supply the monthly power needs of 1,000 typical homes in the Western U.S. (This is a rounding up to 8,760 kWh/year/home based on an average of 8,549 kWh used/household/year. (U.S. DOE EIA, 1997 annual/capita electricity consumption figures.))

METER - A device used to measure some quantity, for example, electrical demand, electrical energy, temperature and flow. A device for measuring levels and volumes of a customer's gas or electricity use.

METERED DATA - Data collected at customer premises over time through a meter for a specific end-use or energy-using system (e.g., lighting and HVAC), or location (e.g., floors of a building or a whole premise). Metered data may be collected over a variety of time intervals. Usually refers to electricity or gas data.

METERED DEMAND - The average time rate of energy flow over a period of time recorded by a utility meter.

METERING - The collection of energy consumption data over time at customer premises through the use of meters. These meters may collect information about kWh, kW or therms, with respect to an end-use, a circuit, a piece of equipment or a whole building (or facility). Short-term metering generally refers to data collection for no more than a few weeks. End-use metering refers specifically to separate data collection for one or more end-uses in a building, such as lighting, air conditioning or refrigeration. What is called “spot metering” is not metering in this sense, but is an instantaneous measurement (rather than over time) of volts, amps, watts or power factor to determine equipment size and/or power draw.

METRIC - A point of measurement. Any point of measurement that can be defined, quantified and assessed.

MODEL - A mathematical representation or calculation procedure that is used to predict the energy use and demand in a building or facility or to estimate efficiency program savings estimates. Models may be based on equations that specifically represent the physical processes or may be the result of statistical analysis of energy use data.

MONITORING (equipment or system) - Gathering of relevant measurement data over time to evaluate equipment or system performance, e.g., chiller electric demand, inlet evaporator temperature and flow, outlet evaporator temperature, condenser inlet temperature, and ambient dry-bulb temperature and relative humidity or wet-bulb temperature, for use in developing a chiller performance map (e.g., kW/ton vs. cooling load and vs. condenser inlet temperature).

MULTICOLLINEARITY - When two or more independent variables in a regression model are highly correlated with each other producing high standard errors for the regression parameter. The mathematics of a regression model fail if there is perfect collinearity, an exact linear relationship between two or more independent variables. If the correlation between independent variables is higher than either has with the dependent variable, the problems of multicollinearity are highly likely.

NAIC - North American Industry Classification.

NATURAL CHANGE - The change in base usage over time. Natural change represents the effects of energy-related decisions that would have been made in the absence of the utility programs by both program participants and non-participants.

NET LOAD IMPACT - The total change in load that is attributable to the utility DSM program. This change in load may include, implicitly or explicitly, the effects of free-drivers, free-riders,

state or federal energy efficiency standards, changes in the level of energy service and natural change effects.

NET-TO-GROSS RATIO (NTGR) - A factor representing net program load impacts divided by gross program load impacts that is applied to gross program load impacts to convert them into net program load impacts. This factor is also sometimes used to convert gross measure costs to net measure costs.

NEW CONSTRUCTION - Residential and nonresidential buildings that have been newly built or have added major additions subject to California Code of Regulation Title 24, the California building standards code.

NON-PARTICIPANT - Any customer who was eligible but did not participate in the utility program under consideration in a given program year. Each evaluation plan should provide a definition of a non-participant as it applies to a specific study.

NONRESIDENTIAL – Used to describe facilities used for business, commercial, agricultural, institutional and industrial purposes.

NONRESIDENTIAL BUILDING - Any building which is heated or cooled in its interior and is of an occupancy type other than Type H, I or J, as defined in the Uniform Building Code, 1973 edition, as adopted by the International Conference of Building Officials.

NORMALIZATION - Adjustment of the results of a model due to changes in baseline assumptions (non-independent variables) during the test or post-retrofit period.

NTGR – See NET-TO-GROSS RATIO.

NULL HYPOTHESIS – See SIGNIFICANCE TEST.

OCCUPANCY SENSOR - A control device that senses the presence of a person in a given space, commonly used to control lighting systems in buildings.

ORIENTATION - The position of a building relative to the points of a compass.

P-VALUE – See PROBABILITY-VALUE.

PARTICIPANT - An individual, household, business or other utility customer that received a service or financial assistance offered through a particular utility program, set of utility programs or particular aspect of a utility program in a given program year. The term “service” is used in this definition to suggest that the service can be a wide variety of services, including financial rebates, technical assistance, product installations, training, energy efficiency information or other services, items or conditions. Each evaluation plan should present the definition of a “participant” as it applies to a specific study.

PARTICIPANT TEST - A cost-effectiveness test intended to measure the cost-effectiveness of energy efficiency programs from the perspective of electric and/or gas customers (individuals or organizations) participating in them.

PARTIES OR INTERESTED PARTIES - Persons and organizations with an interest in energy efficiency that comment on or participate in the CPUC's efforts to develop and implement ratepayer-funded energy efficiency programs.

PEAK DEMAND - The maximum level of metered demand during a specified period, such as a billing month or during a specified peak demand period.

PEAK DEMAND PERIOD - Noon to 7 p.m. Monday through Friday, June, July, August and September.

PEAK LOAD - The highest electrical demand within a particular period of time. Daily electric peaks on weekdays occur in late afternoon and early evening. Annual peaks occur on hot summer days.

PERFORMANCE DEGRADATION - Any over time savings degradation (or increases compared to standard efficiency operation) that includes both (1) technical operational characteristics of the measures, including operating conditions and product design, and (2) human interaction components and behavioral measures.

PERSISTENCE STUDY - A study to assess changes in net program impacts over time (including retention and degradation).

PGC - See PUBLIC GOODS CHARGE.

PORTFOLIO - All IOU and non-IOU energy efficiency programs funded through the PGC that are implemented during a program year or cycle.

POST-BUY ANALYSIS - A comparison of the actual advertising schedule run to the original expectations of the schedule as purchased, considering adherence to buy specifications, actual audience achieved as measured by audience ratings services when available, and conformity to standard industry practices. The term is used primarily in relation to broadcast media (and more frequently performed for TV schedules than for radio), but a similar type of stewardship should be performed for purchases of print and outdoor media as well

POST-RETROFIT PERIOD - The time following a retrofit during which savings are to be determined.

POWER ANALYSIS - A power analysis, executed when a study is being planned, is used to anticipate the likelihood that the study will yield a significant effect and is based on the same factors as the significance test itself. Specifically, the larger the effect size used in the power analysis, the larger the sample size; the larger (more liberal) the criterion required for significance (alpha), the higher the expectation that the study will yield a statistically significant

effect. The probability-value (p-value) provided by the significance test and used to reject the null hypothesis, is a function of three factors: size of the observed effect (e.g., gross energy savings), sample size and the criterion required for significance (alpha, the level of confidence). These three factors, together with power, form a closed system – once any three are established, the fourth is completely determined. The goal of power analysis is to find an appropriate balance among these factors by taking into account the substantive goals of the study and the resources available to the researcher.

PRACTICE - Generally refers to a change in a customer's behavior or procedures that reduces energy use (e.g., thermostat settings and maintenance procedures).

PRACTICE RETENTION STUDY - An assessment of the length of time a customer continues the energy conservation behavioral changes after adoption of these changes.

PRECISION - The indication of the closeness of agreement among repeated measurements of the same physical quantity. In econometrics, the accuracy of an estimator as measured by the inverse of its variance.

PROBABILITY-VALUE (P-VALUE) - The probability of obtaining a finding at least as "impressive" as that obtained, *assuming* the null hypothesis is true, so that the finding was the result of chance alone. The p-value is provided by the significance test and used to reject the null hypothesis, and is a function of three factors: size of the observed effect (e.g., gross energy savings), sample size and the criterion required for significance (alpha, the level of confidence). These three factors, together with power, form a closed system – once any three are established, the fourth is completely determined.

PROCESS EVALUATION - A systematic assessment of an energy efficiency program for the purposes of documenting program operations at the time of the examination, and identifying and recommend improvements that can be made to the program to increase the program's efficiency or effectiveness for acquiring energy resources while maintaining high levels of participant satisfaction.

PROCESS OVERHAUL - Modifications to industrial or agricultural processes to improve their energy use characteristics.

PROGRAM - An activity, strategy or course of action undertaken by an implementer or Administrator using PGC funds. Each program is defined by a unique combination of program strategy, market segment, marketing approach and energy efficiency measure(s) included.

PROGRAM (IMPLEMENTATION) CYCLE - The period of time during which programs are funded, planned and implemented. Can be an annual cycle, a bi-annual cycle or other period of time.

PROGRAM DESIGN - The method or approach for making, doing or accomplishing an objective by means of a program.

PROGRAM DEVELOPMENT - The process by which ideas for new or revised energy efficiency programs are converted into a design to achieve a specific objective.

PROGRAM PENETRATION - The level of program acceptance among qualified customers.

PROGRAM MANAGEMENT - The responsibility and ability to oversee and guide the performance of a program to achieve its objective.

PROGRAM STRATEGIES - Refers to the type of method deployed by the program in order to obtain program participation. Some examples of program strategies include: rebates, codes, performance contracting and audits.

PROGRAM THEORY - A presentation of the goals of a program, incorporated with a detailed presentation of the activities that the program will use to accomplish those goals and the identification of the causal relationships between the activities and the program's effects.

PROGRAM YEAR (PY) - The calendar year approved for program implementation. Note that program years can be shorter than 12 months if programs are approved after the beginning of a calendar year (after January 1 of a given year).

PROGRAMMABLE CONTROLLER - A device that controls the operation of electrical equipment (such as air conditioning units and lights) according to a pre-set time schedule.

PROJECT - An activity or course of action undertaken by an implementer involving one or multiple energy efficiency measures, usually at a single site.

PROJECT DEVELOPMENT - The process by which an implementer identifies a strategy or creates a design to provide energy efficiency products, services and practices directly to customers.

PUBLIC GOODS CHARGE (PGC) (Electric) - Per Assembly Bill (AB) 1890, a universal charge applied to each electric utility Customer's bill to support the provision of public goods. Public goods covered by California's electric PGC include public purpose energy efficiency programs, low-income services, renewables, and energy-related research and development.

RATIO ESTIMATOR (SAMPLING METHOD) - A sampling method to obtain increased precision by taking advantage of the correlation between an auxiliary variable and the variable of interest to reduce the coefficient of variation.

REBATES - A type of incentive provided to encourage the adoption of energy-efficient practices, typically paid after the measure has been installed. There are typically two types of rebates: a Prescriptive Rebate, which is a prescribed financial incentive/unit for a prescribed list of products, and a Customized Rebate, in which the financial incentive is determined using an analysis of the customer's equipment and an agreement on the specific products to be installed. Upstream rebates are financial incentives provided for manufacturing, sales, stocking or other

per unit energy-efficient product movement activities designed to increase use of particular type of products.

REBOUND EFFECT – SEE TAKEBACK EFFECT

REGRESSION MODEL - A mathematical model based on statistical analysis where the dependent variable is regressed on the independent variables which are said to determine its value. In so doing, the relationship between the variables is estimated statistically from the data used.

RELIABILITY - When used in energy evaluation refers to the likelihood that the observations can be replicated.

REMODELING – Modifications to or the act of modifying the characteristics of an existing residential or nonresidential building or energy-using equipment installed within it.

RENEWABLE ENERGY or RENEWABLE RESOURCES or RENEWABLE ENERGY RESOURCES- Renewable energy resources are naturally replenishable, but flow-limited. They are virtually inexhaustible in duration but limited in the amount of energy that is available per unit of time. Some (such as geothermal and biomass) may be stock-limited in that stocks are depleted by use, but on a time scale of decades or perhaps centuries, they can probably be replenished. Renewable energy resources include: biomass, hydro, geothermal, solar and wind. Renewable resources also include some experimental or less-developed sources such as the use of ocean thermal, wave and tidal action technologies. Utility renewable resource applications include bulk electricity generation, on-site electricity generation, distributed electricity generation, non-grid-connected generation and demand-reduction (energy efficiency) technologies.

RENOVATION - Modification to the characteristic(s) of an existing residential or nonresidential building, including but not limited to windows, insulation and other modifications to the building shell.

REPLACEMENT - Refers to the changing of equipment either due to failure, move to more efficient equipment or other reasons near the end of product life or earlier. Often used to refer to a move to a more energy-efficient product that replaces an inefficient product.

RESEARCH AND DEVELOPMENT (R&D) - Research is the discovery of fundamental, new knowledge. Development is the application of new knowledge to develop a potential new service or product. Basic power sector R&D is most commonly funded and conducted through the Department of Energy (DOE), its associated government laboratories, university laboratories, the Electric Power Research Institute (EPRI) and private sector companies.

RESIDENTIAL BUILDING - Means any hotel, motel, apartment house, lodging house, single dwelling or other residential building that is heated or mechanically cooled.

RESIDENTIAL CUSTOMER – Utility customers with accounts for existing single-family residences, multi-family dwellings (whether master-metered or individually metered) and buildings that are essentially residential but used for commercial purposes, including but not limited to time shares and vacation homes.

RETAIL MARKET - A market in which electricity and other energy services are sold directly to the end-use customer.

RETENTION (MEASURE) - The degree to which measures are retained in use after they are installed.

RETROFIT - Energy efficiency activities undertaken in existing residential or nonresidential buildings where existing inefficient equipment is replaced by efficient equipment.

RETROFIT ISOLATION - The savings measurement approach defined in IPMVP Options A and B, and ASHRAE Guideline 14 that determines energy or demand savings through the use of meters to isolate the energy flows for the system(s) under consideration.

RIGOR - The level of expected reliability. The higher the level of rigor, the more confident we are the results of the evaluation are both accurate and precise, i.e., reliable. That is, reliability and rigor are treated as synonymous. Reliability is discussed in the Sampling and Uncertainty Protocol and in the *Evaluation Framework* where it is noted that sampling precision does not equate to accuracy. Both are important components of reliability, as used in this Protocol.

SAE - See STATISTICALLY ADJUSTED ENGINEERING MODELS.

SAMPLE DESIGN - The approach used to select the sample units.

SAMPLING ERROR - An error which arises because the data are collected from a part, rather than the whole of the population. It is usually measurable from the sample data in the case of probability sampling.

SAVINGS MEASUREMENT APPROACH - The estimation of energy and demand savings associated with an energy efficiency measure for a piece of equipment, a subsystem or a system. The estimated savings are based on some kind of measured data from before and after the retrofit and may be calculated using a variety of engineering techniques.

SERIAL CORRELATION - See AUTOCORRELATION.

SERVICE AREA - The geographical territory served by a utility.

SETBACK THERMOSTAT - See THERMOSTAT, SETBACK.

SHADING - The protection from heat gains due to direct solar radiation. Shading is provided by permanently attached exterior devices, glazing materials, and adherent materials applied to the glazing or an adjacent building for nonresidential buildings, hotels, motels and high rise

apartments, and by devices affixed to the structure for residential buildings. (See California Code of Regulations, Title 24, Section 2-5302.)

SHADING COEFFICIENT (SC) - The ratio of solar heat gain through fenestration, with or without integral shading devices, to that occurring through unshaded 1/8 in. thick clear double strength glass. See also **SOLAR HEAT GAIN COEFFICIENT**.

SHGC - See **SOLAR HEAT GAIN COEFFICIENT**.

SIGNIFICANCE TEST – Traditionally, data collected in a research study is submitted to a significance test to assess the viability of the null hypothesis. The null hypothesis is a term that statisticians often use to indicate the statistical hypothesis tested. The purpose of most statistical tests is to determine if the obtained results provide a reason to reject the hypothesis that they are merely a product of chance factors. For example, in an experiment in which two groups of randomly selected subjects have received different treatments and have yielded different means, it is always necessary to ask if the difference between the obtained means is among the differences that would be expected to occur by chance whenever two groups are randomly selected. In this example, the hypothesis tested is that the two samples are from populations with the same mean. Another way to say this is to assert that the investigator tests the null hypothesis that the difference between the means of the populations, from which the samples were drawn, is zero. If the difference between the means of the samples is among those that would occur rarely by chance when the null hypothesis is true, the null hypothesis is rejected and the investigator describes the results as statistically significant.

SIMPLE RANDOM SAMPLING - A method of selecting n sample units out of the N population such that every one of the distinct N items has an equal chance of being selected.

SIMPLIFIED ENGINEERING MODEL - Engineering equations used to calculate energy usage and/or savings. These models are usually based on a quantitative description of physical processes that describe the transformation of delivered energy into useful work such as heat, lighting or motor drive. In practice, these models may be reduced to simple equations that calculate energy usage or savings as a function of measurable attributes of customers, facilities or equipment (e.g., lighting use = watts X hours of use). These models do not incorporate billing data and do not produce estimates of energy savings to which tests of statistical validity can be applied.

SNAPBACK EFFECT – SEE **TAKEBACK EFFECT**

SOLAR HEAT GAIN - Heat added to a space due to transmitted and absorbed solar energy.

SOLAR HEAT GAIN COEFFICIENT (SHGC) - The ratio of the solar heat gain entering the space through the fenestration area to the incident solar radiation.

SOLAR HEAT GAIN FACTOR - An estimate used in calculating cooling loads of the heat gain due to transmitted and absorbed solar energy through 1/8"-thick, clear glass at a specific latitude, time and orientation.

SOLAR HEATING AND HOT WATER SYSTEMS - Solar heating or hot water systems provide two basic functions: (1) capturing the sun's radiant energy, converting it into heat energy and storing this heat in insulated storage tank(s); and (2) delivering the stored energy as needed to either the domestic hot water or heating system. These components are called the collection and delivery subsystems.

SPILLOVER - Reductions in energy consumption and/or demand in a utility's service area caused by the presence of the DSM program, beyond program related gross or net savings of participants. These effects could result from: (a) additional energy efficiency actions that program participants take outside the program as a result of having participated; (b) changes in the array of energy-using equipment that manufacturers, dealers and contractors offer all customers as a result of program availability; and (c) changes in the energy use of non-participants as a result of utility programs, whether direct (e.g., utility program advertising) or indirect (e.g., stocking practices such as (b) above or changes in consumer buying habits). Spillover impacts are to be evaluated via the Impact Evaluation Protocols (participant spillover) or by the Market Effects Protocol (non-participant spillover), but spillover impacts are not to be counted toward goal achievements at this time.

SPURIOUSNESS OR SPURIOUS CORRELATION - The apparent association between two variables that is actually attributable to a third variable outside the current analysis, probably a common precedent variable.

STAKEHOLDERS - In program evaluation, stakeholders refer to the myriad of parties that are impacted by a program. Stakeholders include: regulatory staff, program designers, implementers and evaluators, energy producers, special interest groups, potential participants and customers.

STANDARD DEVIATION - The square root of the variance.

STATEWIDE MARKETING AND OUTREACH PROGRAMS - Programs that convey consistent statewide messages to individual consumers through a mass-market advertising campaign.

STATEWIDE PROGRAM - A program available in the service territories of all four large California IOUs, with identical implementation characteristics in all areas, including incentives and application procedures.

STATISTICAL ANALYSIS - Extrapolation of sample data up to the population, calculation of error bounds.

STATISTICAL COMPARISONS - A comparison group of customers serving as a proxy of what program participants would have looked like if the program had not been offered.

STATISTICAL POWER - The probability that statistical significance will be attained, given that there really is a treatment effect. From Lipsey, Mark W. *Design Sensitivity: Statistical Power for Experimental Research*. Newbury Park, CA: SAGE Publications, 1990.

STATISTICALLY ADJUSTED ENGINEERING (SAE) MODELS - A category of billing analysis models that incorporate the engineering estimate of savings as a dependent variable. The regression coefficient in these models is the percentage of the engineering estimate of savings observed in changes in energy usage. For example, if the coefficient on the SAE term is 0.8, this means that the customers are on average realizing 80 percent of the savings from their engineering estimates.

STRATIFIED RANDOM SAMPLING – A sampling method in which the population is divided into X units of subpopulations that are non-overlapping and together comprise the entire population, called strata. A simple random sample is taken of each strata to create a sample based upon stratified random sampling.

STRATIFIED RATIO ESTIMATION - A sampling method that combines a stratified sample design with a ratio estimator to reduce the coefficient of variation by using the correlation of a known measure for the unit (e.g., expected energy savings) to stratify the population and allocate sample from strata for optimal sampling.

SUPPLY-SIDE - Activities conducted on the utility's side of the customer meter. Activities designed to supply electric power to customers, rather than meeting load through energy efficiency measures or on-site generation on the customer side of the meter.

SURVIVAL ANALYSIS - Survival analysis is a class of statistical methods for studying the timing of events or time-to-event models. Originally these models were developed for medical research where the time to death was analyzed, hence the name survival analysis. These statistical methods are designed to work with time-dependent covariates and censoring. Time dependent covariates are independent variables whose impacts on the dependent variable vary by not only its occurrence but also its timing. Censored data refers to not knowing when something occurred because it is before your data collection (left-censored) or has yet to occur at the time of data collection (right-censored).

SYSTEM - A combination of equipment and/or controls, accessories, interconnecting means and terminal elements by which energy is transformed so as to perform a specific function, such as HVAC, service water heating or illumination.

TAKEBACK EFFECT – A change in energy using behavior that yields an increased level of service and that occurs as a result of taking an energy efficiency action.

TECHNICAL DEGRADATION FACTOR - A multiplier used to account for time- and use-related change in the energy savings of a high efficiency measure or practice relative to a standard efficiency measure or practice due to technical operational characteristics of the measures, including operating conditions and product design.

TECHNICAL POTENTIAL - The complete penetration of all measures analyzed in applications where they were deemed technically feasible from an engineering perspective.

TEMPERATURE - Degree of hotness or coldness measured on one of several arbitrary scales based on some observable phenomenon (such as the expansion).

THERM - One hundred thousand (100,000) British thermal units (1 therm = 100,000 Btu).

THERMOSTAT - An automatic control device designed to be responsive to temperature and typically used to maintain set temperatures by cycling the HVAC system.

THERMOSTAT, SETBACK - A device containing a clock mechanism, which can automatically change the inside temperature maintained by the HVAC system according to a pre-set schedule. The heating or cooling requirements can be reduced when a building is unoccupied or when occupants are asleep. (See California Code of Regulations, Title 24, Section 2- 5352(h).)

TIME-OF-USE (TOU) METER - A measuring device that records the times during which a customer uses various amounts of electricity. This type of meter is used for customers who pay time-of-use rates.

TIME-OF-USE (TOU) RATES - Electricity prices that vary depending on the time periods in which the energy is consumed. In a time-of-use rate structure, higher prices are charged during utility peak-load times. Such rates can provide an incentive for consumers to curb power use during peak times.

TOTAL FLOOR AREA is the floor area (in square feet) of enclosed space on all floors of a building, as measured at the floor level of the exterior surfaces of exterior walls enclosing the space.

TOTAL RESOURCE COST TEST – SOCIETAL VERSION - A cost-effectiveness test intended to measure the overall cost-effectiveness of energy efficiency programs from a societal perspective.

TOU – See **TIME-OF-USE METER** and **TIME-OF-USE RATES**.

TRIANGULATION - Comparing the results from two or more different data gathering or measurement techniques on the same problem to derive a “best” estimate from the analysis of the comparison.

UA - A measure of the amount of heat that would be transferred through a given surface or enclosure (such as a building envelope) with a 1° F temperature difference between the two sides. The UA is calculated by multiplying the U-value by the area of the surface (or surfaces).

UNCERTAINTY - The range or interval of doubt surrounding a measured or calculated value within which the true value is expected to fall within some degree of confidence.

UNCERTAINTY ANALYSIS - (a) A procedure or method by which the uncertainty of a measured or calculated value is determined; (b) the process of determining the degree of confidence in the true value when using a measurement procedure(s) and/or calculation(s).

UNCONDITIONED SPACE - A space that is neither directly nor indirectly conditioned space, which can be isolated from conditioned space by partitions and/or closeable doors. (See California Code of Regulations, Title 24, Section 2-5302.)

UPGRADE (Electric utility) - Replacement or addition of electrical equipment resulting in increased generation or transmission capability.

UPSTREAM PROGRAMS - Programs that provide information and/or financial assistance to entities in the delivery chain of high-efficiency products at the retail, wholesale or manufacturing level.

UTILITY METER - The meter used to calculate a monthly energy and/or demand charge at a specific utility/customer connection; more than one may be installed per customer and per site due to different supply voltages, capacity requirements, physical separation distances, installation periods or for specific customer requirements or utility programs.

U-VALUE or U-FACTOR - A measure of how well heat is transferred by the entire window - the frame, sash and glass - either into or out of the building. U-value is the opposite of R-value. The lower the U-factor number, the better the window will keep heat inside a home on a cold day.

VALIDITY - The extent to which any measuring instrument measures what it is intended to measure.

VENTILATION - The process of supplying or removing air by natural or mechanical means to or from any space. Such air may or may not have been conditioned or treated.

WATT - A unit of measure of electric power at a point in time, as capacity or demand. One watt of power maintained over time is equal to one joule/second. Some Christmas tree lights use one watt. The watt is named after Scottish inventor James Watt and is capitalized when shortened to W and used with other abbreviations, as in kWh.

WATT-HOUR - One watt of power expended for one hour. One thousandth of a kilowatt-hour.

WEATHERSTRIPPING - Specially designed strips, seals and gaskets installed around doors and windows to limit air leakage.

WET-BULB TEMPERATURE - The temperature at which water, by evaporating into air, can bring the air to saturation at the same temperature. Wet-bulb temperature is measured by a wet-bulb psychrometer.

WHOLE-BUILDING CALIBRATED SIMULATION APPROACH - The savings measurement approach defined in IPMVP Option D and ASHRAE Guideline 14, which involves the use of an approved computer simulation program to develop a physical model of the building in order to determine energy and demand savings. The simulation program is used to model the energy used

by the facility before and after the retrofit. The pre- or post-retrofit models are developed by calibration with measured energy use and demand data and weather data.

WHOLE-BUILDING METERED APPROACH - The savings measurement approach defined in IPMVP Option C and ASHRAE Guideline 14 that determines energy and demand savings through the use of whole-facility energy (end-use) data, which may be measured by utility meters or data loggers. This approach may involve the use of monthly utility billing data or data gathered more frequently from a main meter.

ZONE - A space or group of spaces within a building with any combination of heating, cooling or lighting requirements sufficiently similar so that desired conditions can be maintained throughout by a single controlling device.

APPENDIX C. Performance Basis Metrics

This section identifies the performance basis metrics that the CPUC-ED will use to assess the performance of the Administrator's energy efficiency programs. These metrics are to be provided by the Administrators to the CPUC-ED.

When an evaluation plan indicates that one or more of the following metrics will be collected, assessed or reported within an evaluation effort, the evaluation contractor will report the performance basis metrics for each program being evaluated and for the aggregation of programs when the evaluation includes more than one program. The performance basis metrics are to be reported in an appendix to the evaluation plan entitled Performance Basis Metric Reporting. The evaluation contractor is to work with the Joint Staff during the evaluation planning efforts to identify the performance basis metrics that are to be reported within the evaluation effort.

The following is a list of the performance basis metrics that are reported by the program administrators to be considered for inclusion in the evaluation reports. The decision of which metrics to include in the evaluation reports will be made by the Joint Staff and provided to the evaluation contractor. The evaluation contractor is to coordinate with Joint Staff to assure that the appropriate performance basis metrics are included in the evaluation reports.

1. Measure installation counts reported by the program.
2. Program costs reported by the program.
3. Measure-specific unit Energy Savings reported by the program.
4. Measure-specific installations by program delivery strategy reported by the program.
5. Program administrator estimates of Gross Energy Savings.
6. Program administrator estimated net-to-gross ratios by measure and delivery strategy.
7. Program administrator estimates of net energy savings.
8. Load factors or daily load shapes used to transform annual savings estimates into peak
9. savings estimates.
10. Incremental measure costs.

Appendix D. A Primer for Using Power Analysis to Determine Sample Sizes

Power is the probability that you will detect an “effect” that is there in the true population that you are studying. Put another way, the power of a statistical test of a null hypothesis is the probability that it will lead to a rejection of the null hypothesis when it is false, i.e., the probability that it will result in the conclusion that the phenomenon exists. The “effect” could be a difference between two means, a correlation between two variables (r), a regression coefficient (b), a chi-squared, etc. Power analysis is a statistical technique that can be used (among other things) to determine sample size requirements to ensure that statistical significance can be found. This appendix provides an overview of using power analysis for determining required sample sizes. It provides references and an example of using power analysis for this purpose.

Power analysis is a required component in several of the Energy Efficiency Evaluation Protocols to assist in determining required sample sizes for the allowable methods that use any type of regression analysis. The regression-based methods within the Impact Protocol¹⁴² and the Effective Useful Life Evaluation Protocol (Retention and Degradation) (e.g., survival analysis) must use power analysis to plan their sample size (unless census samples are being used). (Regression-based methods must also meet the requirements of the Sampling and Uncertainty Protocol.)

In all of the Protocols, where power analysis is required it is one of up to three inputs to be used for the determination of sample size for a non-census regression study. Each Protocol states that power analysis, results from prior studies on similar programs, and professional judgment are to be used to determine the required sample size. Sample size planning is an important component in the evaluation planning activity. The proposed sample size(s) must be within the evaluation plan submitted and approved by Joint Staff prior to undertaking sample data collection.

There are many possible references for power analysis and over the last decade it has become a standard component of graduate statistics courses. The seminal work was conducted by Jacob Cohen and the classic text cited is his 1988 *Statistical Power Analysis for the Behavioral Sciences*. Power analysis can be used for many things but is only being required in the Protocol for determining required sample sizes.

There are several software packages and calculation Web sites that conduct the power analysis calculation. The National Institute of Health provided funding to BioStat, Inc. to create stand-alone software to conduct power analysis calculations. The current version of this software is called Power and PrecisionTM and is offered for sale by BiostatTM (www.PowerAnalysis.com). The major statistical software packages that evaluators are likely to use for conducting regression-based analyses have incorporated components that conduct power analysis. For

¹⁴² These include the Gross Energy Impact Protocol, Gross Demand Impact Protocol, Participant Net Impact Protocol, and the Indirect Impact Protocol. All of these have at least one minimum allowable method that is regression-based. Regression-based methods discussed in these Protocols include, but are not limited to, multiple regression (econometric analysis), Analysis of Covariance (ANCOVA), and discrete choice (logistic regression).

example, it is included in SPSS® and in the 9.1 versions of SAS® along with the SAS/STAT® package (Power and Precision module).

A brief overview of the parameters to be input for conducting power analysis for the purpose of determining required sample size for the primary regression model types primarily used within energy efficiency evaluation is provided below. This is followed by an example where power analysis is conducted to determine the required sample size for a survival analysis (the preferred methodology for effective useful life analysis). This example illustrates how the sample size requirement varies according to different input parameters.

A small list of references is provided at the end of this Appendix.

Basics of Power Analysis and the Protocols

There is some variation in the parameters and the set-up required to use power analysis to determine sample requirements for different types of analyses. There are some that are common to all power analysis. There are four common parameters that create a closed system for power analysis. These are:

- Alpha
- Power
- Effect size
- Sample size

Alpha is the criterion required to establish statistical significance. For consistency across studies, these Protocols have called for 90% confidence level (precision) and then varied the error tolerance based upon the rigor level assigned. A 90% precision equates to an alpha of 0.10. This represents the probability or proportion of studies that would result in a Type I error, where the researcher rejects the null hypothesis when it is in fact true. For consistency with the precision requirements elsewhere in the Protocols, the alpha should be set at 0.10 when using power analysis to determine the required sample size.

Power is the probability that one find a statistically significant effect (when in reality there is one), assuming the effect size, sample size, and alpha criteria. It is common to set power from 0.7 to 0.9. The EUL Analysis Protocol sets the minimum power to 0.7 for the Basic rigor level and 0.8 for the Enhanced rigor.

The effect size is the expected magnitude of the effect. However, effect size will be expressed differently depending on the unit of measurement of the variables involved and on the type of analysis being performed.

In determining sample sizes in the research planning process, values for these parameters can be varied in an attempt to balance a level of statistical power, the alpha, and the effect size, all determined with an eye on the budget constraints. In the end, the results of the power analysis will be combined with professional judgment and past studies to arrive at the required sample sizes.

For multiple regression, analysis of covariance (ANCOVA), and logistic regression, there are three parameters that one can vary when determining the required sample size:

- Alpha
- Power
- Effect size

For survival analysis, there is a fourth parameter that can be varied, the duration of study. Survival analysis depends upon failures to estimate the function of when failure will occur taking into account that for many of the sites failures will not have yet been observed (i.e., the data is right-censored, the point of failure is not determined for many in the sample). The later the study (i.e., the greater the duration), the greater the power since a greater duration increases probability that more failures will be observed. For the same alpha, effect size, and power, a study that plans to collect retention data close to the ex-ante EUL (the median measure life) will require fewer sample points than a study conducted earlier.

We conclude this brief introduction with a list of power facts.

- The more stringent the significance level, the greater the necessary sample size. More subjects are needed for a 1% level test than for a 5% level test.
- Two-tailed tests require larger sample sizes than one-tailed tests. Assessing two directions at the same time requires a greater investment. (At the same time, good science requires that a one-tail test is only used when there is strong proof that it is appropriate to do so and not being used for the purpose of making it simpler to pass a statistical significance test.)
- The smaller the critical effect size, the larger the sample size. Subtle effects require greater efforts.
- The larger the power required, the larger the necessary sample size. Greater protection from failure requires greater effort.
- The smaller the sample size, the smaller the power, i.e., the greater the chance of failure.
- If one proposed to conduct an analysis with a very small sample size, one must be willing to accept a high risk of finding no statistically significant results, or be operating in an area in which the critical effect size is quite large.

Example of Varying Parameters and Estimating Required Sample Size for Survival Analysis through Power Analysis

The basic level of rigor in the EUL Protocols requires that a 0.70 level of power be planned at the 90% level of confidence. While the enhanced level of rigor requires that a 0.80 level of power be planned also at the 90% level of confidence. An exercise was conducted using Power and PrecisionTM software to provide an example of the use of power analysis to set required

sample size and to demonstrate the impact of the different power level requirements on sample size requirements.¹⁴³

Two hypothetical situations were created around an energy efficiency measure with an ex-ante median EUL of 8 years.

- In the first situation, a researcher is interested in setting up a study to detect an effect size (a delta) of two years in both directions. In other words, our ex-post estimate around an 8 year median EUL finding would be 6 years to 10 years.
- In the second situation, a researcher is interested in setting up a study to detect an effect size (a delta) of only one year in both directions. In other words, our ex-post estimate around an 8 year median EUL finding would be 7 years to 9 years.

In both cases it was assumed that the effect was selected as the smallest effect that would be important to detect, in the sense that any smaller effect would not be of any substantive significance. It is also assumed that the effect size is reasonable, in the sense that an effect of this magnitude could be anticipated in this field of research.

The conditions of the study were as follows:

- A two-tailed test was used since it is possible that the ex-post EUL could be higher or lower than the ex-ante EUL.
- The computation assumes an attrition rate of zero. This means that all sites will be followed until the measure is no longer operational or is not there (the terminal event) or until the study ends.
- This study assumes a condition in which subjects are entered during a given program period and then followed until either (a) the terminal event occurs, or (b) the study ends causing us not to know how long the equipment will last in all those sites that still have operational equipment (i.e., the site is randomly right-censored). The study design calls for all subjects to be accrued before the study begins, with the retention study to occur at 5 years after the program year under study (a follow-up period of 5 years). In other words, all subjects in the sample will be followed for a maximum of 5 years.
- The alpha level was set at 0.10. (This equates to 90% precision.)
- This study systematically varied two levels of power (0.70 and 0.80) to examine the impacts of varying the required power on the subsequent required sample size.
- Finally, this study systematically varied two levels of effect size to examine the impact of alternative effect size requirements. Both assumed that the ex-ante EUL was 8 years. For the ex-post EUL we first assumed 10 years, which means that the delta between ex-ante and ex-post is two years. We then assumed 9 years, which means that the delta between ex-ante and ex-post is only one year, a much smaller effect.

Table 25 below shows the effect on sample sizes of varying both the effect and the power. The differences looking from one column to the other column demonstrate the differential impact of

¹⁴³ Produced by Biostat™. Information available at: www.PowerAnalysis.com

requiring a power of 0.7 versus 0.8. Looking from row to row demonstrates the impact on sample size requirements of desiring to obtain a one-year effect versus a two-year effect (for the 8 year ex-ante survival analysis).

Table 25. Sample Sizes as a Function of Alpha and Power

Effect	Power	
	0.70	0.80
1 Year	1,050	1,400
2 Years	320	420

As one can see, at alphas of 0.70 and 0.80, the sample sizes increase by approximately 33 percent for a one-year effect size and 31.3 percent for a two-year effect size. Increasing the power requirement from 0.70 to 0.80 increases the required sample size by approximately one-third. However, as one moves from the effect of one year to two years, the required sample sizes increase by approximately 230 percent, from 320 to 1,050 for a power of 0.7 and from 420 to 1,400 for a power of 0.8. Clearly, the impact of a smaller effect is greater than the impact of increasing the power.

While we have attempted to keep the example simple so that the effect of moving from the standard to the enhanced level of rigor can be clearly seen, we note that there are four parameters that one can adjust for determining the required sample size. These are the:

- Duration of study (the post 5-year study assumption in our example.)¹⁴⁴
- Alpha (The precision level which we set at 90% confidence, as is done throughout the evaluation Protocols, which provides an alpha of 0.1.)
- Power
- Effect size

Consider the case in which the effect is one year at a power of 0.80, requiring a sample size of 1,400. If one chose an alpha of 0.20 (as was done in the pre-1998 Protocols for the EUL analysis) and extended the follow-up period from 5 years to 7 years, then the sample size is reduced to 770.

In determining sample sizes in the research planning process, values for these parameters can be varied in an attempt to balance a level of statistical power, the alpha, the duration of the study, and the effect size, all determined with an eye on the budget constraints. The values will probably be unique to each measure selected for study. In the end, the results of the power analysis will be combined with professional judgment and past studies to arrive at the required sample sizes.

¹⁴⁴ The closer the study is to the ex-ante EUL the lower the sample size requirement since finding enough failures to complete the analysis is a primary component of sample size requirement and the ability to obtain a survival analysis result.

References

Borenstein, Michael, and Jacob Cohen. (1988) *Statistical Power Analysis: A Computer Program*. Hillsdale, N.J.: Lawrence Erlbaum Associates, Publishers.

Borenstein, Michael, Hannah Rothstein, and Jacob Cohen. (1988) *Power and PrecisionTM*. Englewood, N.J.: Biostat, Inc.

Cohen, Jacob. (1988) *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, N.J.: Lawrence Erlbaum Associates, Publishers.

Kraemer, Helena Chmura and Sue Thiemann. (1987) *How Many Subjects? Statistical Power Analysis in Research*. Newbury Park, CA: SAGE Publications.

Lipsey, Mark W. (1990) *Design Sensitivity: Statistical Power for Experimental Research*. Newbury Park, CA: SAGE Publications.

Statistical Package for the Social Sciences, SPSS[®] (See SPSS.com)

Statistical Analysis Software (SAS[®])

Appendix E. Summary Tables for All Protocols

The following tables are provided as a quick reference to the summary tables found in the Protocols.

Summary of Protocol-Driven Impact Evaluation Activities

Required Protocols for Gross Energy Evaluation	
Rigor Level	Minimum Allowable Methods for Gross Energy Evaluation
Basic	1. Simple Engineering Model (SEM) with M&V equal to IPMVP Option A and meeting all requirements in the M&V Protocol for this method. Sampling according to the Sampling and Uncertainty Protocol. 2. Normalized Annual Consumption (NAC) using pre- and post-program participation consumption from utility bills from the appropriate meters related to the measures undertaken, normalized for weather, using identified weather data to normalize for heating and/or cooling as is appropriate to measures included. Twelve (12) months pre-retrofit and twelve (12) months post-retrofit consumption data is required. Sampling must be according to the Sampling and Uncertainty Protocol.
Enhanced	1. A fully specified regression analysis of consumption information from utility bills with inclusion/adjustment for changes and background variables over the time period of analysis that could potentially be correlated with the gross energy savings being measured. Twelve (12) months post-retrofit consumption data are required. Twelve (12) months pre-retrofit consumption data are required, unless program design does not allow pre-retrofit billing data, such as in new construction. In these cases, well-matched control groups and post-retrofit consumption analysis is allowable.¹⁴⁵ Sampling must be according to the Sampling and Uncertainty Protocol utilizing power analysis as an input to determining required sample size(s). 2. Building energy simulation models that are calibrated as described in IPMVP Option D requirements in the M&V Protocols. If appropriate, may alternatively use a process-engineering model (e.g., AirMaster+) with calibration as described in the M&V Protocols. Sampling according to the Sampling and Uncertainty Protocol. 3. Retrofit Isolation engineering models as described in IPMVP Option B requirements in the M&V Protocols. Sampling according to the Sampling and Uncertainty Protocol. 4. Experimental design established within the program implementation process, designed to obtain reliable net energy savings based upon differences between energy consumption between treatment and non-treatment groups from consumption data.¹⁴⁶ Sampling must be according to the Sampling and Uncertainty Protocol.

¹⁴⁵ Post-retrofit only billing collapses the analysis from cross-sectional time-series to cross-sectional. Given this, even more care and examination is expected with regard to controlling for cross-sectional issues that could potentially bias the savings estimate.

¹⁴⁶ The overall goal of the Direct Impact Protocols is to obtain reliable net energy and demand savings estimates. If the methodology directly estimates net savings at the same or better rigor than the required level of rigor, then a gross savings and participant net impact analysis is not required to be shown separately.

Required Protocols for Gross Demand Evaluation	
Rigor Level	Minimum Allowable Methods for Gross Demand Evaluation
Basic	Reliance upon secondary data for estimating demand impacts as a function of energy savings. End-use savings load shapes or end-use load shapes from one of the following will be used to estimate demand impacts: 1. End-use savings load shapes, end-use load shapes or allocation factors from simulations conducted for DEER 2. Allocation factors from CEC forecasting models or utility forecasting models with approval through the evaluation plan review process 3. Allocation based on end-use savings load shapes or end-use load shapes from other studies for related programs/similar markets with approval through the evaluation plan review process
Enhanced	Primary demand impact data must be collected during the peak hour during the peak month for each utility system peak. Estimation of demand impact estimates based on these data is required. If the methodology and data used can readily provide 8,760-hour output, these should also be provided.¹⁴⁷ Sampling requirements can be met at the program level but reporting must be by climate zone (according to CEC's climate zone classification). 1. If interval or time-of-use consumption data are available for participants through utility bills, these data can be used for regression analysis, accounting for weather, day type and other pertinent change variables, to determine demand impact estimates. Pre- and post-retrofit billing periods must contain peak periods. Requires using power analysis, evaluations of similar programs, and professional judgment to determine sample size requirements for planning the evaluation. Needs to meet the requirements of the Sampling and Uncertainty Protocol. 2. Spot or continuous metering/measurement of peak pre and post-retrofit during the peak hour of the peak month for the utility system peak to be used with full measurement Option B or calibrated engineering model Option D meeting all requirements as provided in the M&V Protocol. Pre-retrofit data must be adjusted for weather and other pertinent change variables. Must meet the Sampling and Uncertainty Protocol with a program target of 10% precision at a 90% confidence level. 3. Experimental design established within the program implementation process, designed to obtain reliable net demand savings based upon differences between energy consumption during peak demand periods between treatment and non-treatment groups from consumption data or spot or continuous metering.¹⁴⁸ Sampling must be according to the Sampling and Uncertainty Protocol.

¹⁴⁷ This includes the use of 15-minute interval data or Building Energy Simulation models whose output is 8,760 hourly data.

¹⁴⁸ The overall goal of the Impact Protocols is to obtain reliable net energy and demand savings estimates. If the methodology directly estimates net savings at the same or better rigor than the required level of rigor, then a gross savings and participant net impact analysis is not required to be shown separately.

Required Protocols for Participant Net Impact Evaluation	
Rigor Level	Minimum Allowable Methods for Participant Net Impact Evaluation
Basic	1. Participant self-report.
Standard	- Participant and non-participant analysis of utility consumption data that addresses the issue of self-selection. - Enhanced self-report method using other data sources relevant to the decision to install/adopt. These could include, for example, record/business policy and paper review, examination of other similar decisions, interviews with multiple actors at end-user, interviews with mid-stream and upstream market actors, Title 24 review of typically built buildings by builders and/or stocking practices. - Econometric or discrete choice¹⁴⁹ with participant and non-participant comparison addressing the issue of self-selection.
Enhanced	1. "Triangulation" using more than one of the methods in the Standard Rigor Level. This must include analysis and justification for the method for deriving the triangulation estimate from the estimates obtained.

Required Protocols for Indirect Impact Evaluation	
Rigor Level	Minimum Allowable Methods for Indirect Impact Evaluation
Basic	An evaluation to estimate the program's net changes on the behavior of the participants is required; the impact of the program on participant behavior.
Standard	A two-stage analysis is required that will produce energy and demand savings. The first stage is to conduct an evaluation to estimate the program's net changes on the behavior of the participants/targeted-customers. The second is to link the behaviors identified to estimates of energy and demand savings based upon prior studies (as approved through the evaluation planning or evaluation review process).
Enhanced	A three-stage analysis is required that will produce energy and demand savings. The first stage is to conduct an evaluation to estimate the program's net impact on the behavior changes of the participants. The second stage is to link the behavioral changes to estimates of energy and demand savings based upon prior studies (as approved through the evaluation planning or evaluation review process). The third stage is to conduct field observation/testing to verify that the occurrence of the level of net behavioral changes.

Summary of Protocol-Driven Impact Evaluation Activities	
1	The Joint Staff identifies which programs and program components will receive an impact evaluation and identify the type of impact evaluation(s) to be conducted and at what rigor level.
2	The Joint Staff determines any special needs on a case-by-case basis that will be required from particular program or program component evaluations. CPUC-ED issues request for proposals for

¹⁴⁹ The instrumental-decomposition (ID) method described and referenced in the *Evaluation Framework* (page 145) is an allowable method that falls into this category. A propensity score methodology is also an allowable method in this category as described in: Itzhak Yanovitzky, Elaine Zanutto and Robert Hornik,, "Estimating causal effects of public health education campaigns using propensity score methodology." *Evaluation and Program Planning* 28 (2005): 209–220.

	impact evaluations, selects evaluation contractors and establishes scope(s) of work.
3	Program theory and logic models (PT/LM), if available, must be reviewed/assessed as needed to properly identify impacts and evaluation elements required to assess net program impacts. Research design and sampling plan developed to meet Protocol requirements at a program or program component basis as designated by the Joint Staff rigor level assignments. This includes meeting requirements from the Sampling and Uncertainty Protocol, M&V Protocol and Reporting Protocol, as are applicable given Impact Evaluation Protocol requirements. Research design and sampling must be designed to meet any of the Joint Staff requirements for additional analyses including, but not limited to, the estimation of net impacts by delivery mechanism, the estimation of transmission and/or distribution benefits, or other areas designated of specific concern by the Joint Staff. Develop Evaluation Plan, submit it to the CPUC-ED and revise as necessary to have an approved Evaluation Plan that meets the Impact Evaluation Protocols.
4	All impact evaluation teams must be staffed so as to meet the skills required for the research design, sampling, appropriate and selected impact evaluation method, uncertainty analysis, and reporting being planned and conducted.
5	Develop precise definitions of participants, non-participants and comparison groups. Obtain concurrence with the CPUC-ED on these definitions which are to be used in developing the research design and sampling plans.
6	All impact evaluations must meet the requirements of the Sampling and Uncertainty Protocol. 6.e There are 2 primary sampling considerations for regression-based consumption analysis. (3) Unless a census is utilized, conduct a power analysis to estimate the required sample size. One may also consider prior evaluations for similar programs and professional judgment (must use all of these for the Enhanced level of rigor); and (4) Must use a minimum of 12 months pre and post-retrofit consumption data, except when program approach does not allow pre-retrofit data (e.g., new construction). 6.f All engineering-based methods must: (3) Estimate the uncertainty in all deemed and measured input parameters and consider propagation of error when determining measured quantities and sample sizes to meet the required error tolerance levels; and (4) Use a combination of deemed and measured data sources with sufficient sample sizes designed to meet a 30% error tolerance level in the reported value at a 90% confidence level to meet the Basic rigor level and a 10% error tolerance level at a 90% confidence level for the Enhanced rigor level. 6.g Participant and non-participant comparisons and econometric/discrete-choice methods for Participant Net Impact evaluation will use power analysis combined with examinations of prior evaluation studies for similar programs to derive required sample sizes. 6.h Self-report and Enhanced self-report methods for Participant Net Impact evaluations must at a program level have a minimum sample size of 300 participant decision-makers for at least 300 participant sites (where decision-makers may cover more than one site) or a census attempt, whichever is smaller, (while investigation will be at a measure or end-use level).

7	All impact evaluations must be planned, conducted, analyzed and reported to minimize potential bias in the estimates, justify the methods selected for doing this and report all analysis of potential bias issues as described in the Sampling and Uncertainty Protocol, Impact Evaluation Protocol and M&V Protocol. Primary considerations that must be addressed (based upon method employed) are as follows: 7.e Regression-based consumption analysis must incorporate: (6) Addressing the influence of weather when weather sensitive measures have been included in the program evaluation; (7) Assessing potential bias given inclusion/exclusion issues due to the 12 month pre- and post-retrofit consumption minimum requirement; (8) For the Enhanced rigor level, assess, plan, measure and incorporate background and change variables that might be expected to be correlated with gross and net energy and/or demand savings; (9) Comparison groups must be carefully selected with justification of the criteria for selection of the comparison group and discussion of any potential bias and how the selected comparison group provides the best available minimization of any potential bias; and (10) Interval or TOU consumption data for demand impact analysis must contain the peak period for the utility system peak. If demand billing data is used for demand impact analysis, the research design must address the issues of building demand versus time period for peak and issues with demand ratchets and how the evaluation can reliably provide demand savings estimates. Demand savings must be reported by CTZ. 7.f Engineering-based methods must incorporate: (4) Addressing the influence of weather when weather sensitive measures have been included in the program evaluation; (5) Meeting all the requirements in the M&V Protocol including issues of baseline determination; and (6) For the Enhanced rigor level of demand impact analysis using spot or continuous metering/measurement pre- and post-retrofit for the peak hour of the peak month for the utility system peak. Demand savings must be reported by CTZ. 7.g Experimental design must use spot or continuous metering/measurement pre and post-retrofit for the peak hour of the peak month for the utility system peak for determining demand impacts. Demand savings must be reported by CTZ. 7.h Indirect impact analysis must incorporate: (6) Description of expected impacts (direct behavioral and indirect energy and demand impacts) and how they will be measured; (7) Discussion of identification and measurement of baseline; (8) Extent of exposure/treatment and its measurement;

	(9) Comparison groups must be carefully selected with justification of the criteria for selection of the comparison group and discussion of any potential issues of bias and how the selected comparison group provides the best available minimization of potential bias; and (10) Assessing, planning for and analyzing to control for self-selection bias.
8	Regression analysis of consumption data must address outliers, missing data, weather adjustment, selection bias, background variables, data screens, autocorrelation, truncation, error in measuring variables, model specification and omitted variable error, heteroscedasticity, collinearity and influential data points. These areas must be addressed and reported in accordance with the Sampling and Uncertainty Protocol.
9	Engineering analysis and M&V based methods are required to address sources of uncertainty in parameters, construction of baseline, guarding against measurement error, site selection and non-response bias, engineering model bias, modeler bias, deemed parameter bias, meter bias, sensor placement bias and non-random selection of equipment or circuits to monitor. These areas must be addressed and reported in accordance with the Sampling and Uncertainty Protocol.
10	Develop draft evaluation report to include meeting all requirements in the Reporting Protocol and incorporating the program's performance metrics.
11	Develop final evaluation report in accordance with guidance provided by the Joint Staff. Submit final evaluation report to the CPUC-ED.
12	Once accepted by the CPUC-ED, develop abstracts and post them and report on CALMAC Web site following the CALMAC posting instructions.

Note: The steps included in this evaluation summary table must comply with all the requirements within the Impact Evaluation Protocol.

Summary of Protocol-Driven M&V Activities

Summary of M&V Protocol for Basic Level of Rigor	
Provision	Requirement
Verification	Physical inspection of installation to verify correct measure installation and installation quality
IPMVP Option	Option A ¹⁵⁰
Source of Stipulated Data	DEER assumptions, program work papers, engineering references, manufacturers catalog data, on-site survey data
Baseline Definition	Consistent with program baseline definition. May include federal or Title 20 appliance standards effective at date of equipment manufacture, Title 24 building standards in effect at time of building permit; existing equipment conditions or common replacement or design practices as defined by the program
Monitoring Strategy and Duration	Spot or short-term measurements depending on measure type
Weather Adjustments	Weather dependent measures: normalize to long-term average weather data as directed by the Impact Evaluation Protocol
Calibration Criteria	Not applicable
Additional Provisions	None

Summary of M&V Protocol for Enhanced Level of Rigor	
Provision	Requirement
Verification	Physical inspection of installation to verify correct measure installation and installation quality. Review of commissioning reports or functional performance testing to verify correct operation
IPMVP Option	Option B or Option D
Source of Stipulated Data	DEER assumptions, program work papers, engineering references, manufacturers catalog data, on-site survey data
Baseline Definition	Consistent with program baseline definition. May include federal or Title 20 appliance standards effective at date of equipment manufacture, Title 24 building standards in effect at time of building permit; existing equipment conditions or common replacement or design practices as defined by the program
Monitoring Duration	Sufficient to capture all operational modes and seasons
Weather Adjustments	Weather dependent measures: normalize to long-term average weather data as directed by the Impact Evaluation Protocol
Calibration Criteria	Option D building energy simulation models calibrated to monthly billing or interval demand data. Optional calibration to end-use metered data
Additional Provisions	Hourly building energy simulation program compliant with ASHRAE Standard 140-2001

¹⁵⁰ Exceptions to this provision are programs offering comprehensive measure packages with significant measure interactions; commissioning, and retrocommissioning programs; and new construction programs. Evaluation of measure savings within these programs conducted using engineering methods must follow the Enhanced rigor M&V Protocol and use building energy simulation modeling under IPMVP Option D.

Summary of Protocol-Driven M&V Activities	
1	Receive input from impact evaluation plan. Receive M&V site selection and expected rigor level from the impact evaluation plan.
2	Develop overall M&V plan. The M&V option for each site shall be established according to the rigor assignment and allowable options under the Impact Evaluation Protocol. Project baseline definition with justification shall be reported. Overall M&V planning shall consider the needs of process evaluation studies for measure installation verification and measure performance information. The overall M&V plan shall be submitted for approval to the evaluation project manager as designated by the CPUC-ED.
3	Assess data sources. For each sampled site, the data resources for the engineering analysis must be identified and reviewed. Data sources may include program descriptions, program databases, DEER estimates and underlying documentation, program work papers and on-site surveys. Uncertainties associated with engineering parameters must be estimated. Baseline uncertainties, where not explicitly documented elsewhere, may be informed by professional judgment.
4	Conduct uncertainty analysis. The uncertainty in the estimated savings must be estimated using a propagation of error analysis. The parameters having the greatest influence on the uncertainty must be identified from the propagation of error analysis.
5	Develop site-specific M&V plan according to the outline in the M&V Protocols. The M&V plan must address data collection conducted to reduce uncertainty in the engineering estimates of savings. Sampling of measures within a particular site shall be done in accordance with the Sampling and Uncertainty Protocol. The site-specific M&V plan shall be submitted for review and approval to the evaluation project manager designated by the CPUC-ED prior to commencing field data collection.
6	Conduct pre- and/or post-installation monitoring as indicated by M&V plan. Data collection must be conducted in accordance with the site-specific M&V plan. Changes to the M&V plan resulting from unanticipated field conditions shall be documented and submitted to the evaluation project manager designated by the CPUC-ED.
7	Conduct data analysis and estimate site-specific savings. Conduct analysis of field data and estimate site savings in accordance with site-specific M&V plan. Energy savings estimates for weather-dependent measures shall be normalized to long-term average weather conditions as directed by the Impact Evaluation Protocol.
8	Prepare site-specific M&V report. Prepare a site-specific M&V report for each site used in the analysis that includes the site-specific M&V plan, data collection, data analysis, calculation of measured engineering parameters and overall savings estimates. Calculate the uncertainties associated with energy savings estimates and measurement-derived engineering parameters. The site-specific uncertainty analysis shall include an estimate of the sampling error associated with individual measure sampling within the site, measurement error associated with field data collection and uncertainties associated with any non-measured (deemed) parameters. Potential sources of bias associated with the measurements and engineering analysis shall be identified and steps to minimize the bias shall be reported in accordance with the Sampling and Uncertainty Protocol.
9	Prepare draft overall M&V report. A draft overall M&V project report shall be submitted to the CPUC-ED that meets all the requirements of the Reporting Protocol, demonstrates compliance with the overall M&V plan developed in step 2 and summarizes the results from each site. Site-specific M&V reports shall be included as an Appendix. Raw field data and data analysis results shall be supplied electronically in accordance with the Reporting Protocol.
10	Prepare final overall M&V report. Prepare final overall M&V report in accordance with review comments provided by the Joint Staff.
11	Submit final M&V report. Submit final M&V report and associated datasets to the CPUC-ED.
12	Post final M&V report on the CALMAC Web site. Once accepted by the CPUC-ED, develop abstracts and post them and final M&V report on the CALMAC Web site following the CALMAC posting instructions.

Emerging Technology

Summary of Protocol-Driven Emerging Technology Evaluation Activities	
1	Joint staff selects an evaluation contractor to implement the Emerging Technology Program evaluation.
2	The ETP managers, in collaboration with the evaluation contractor and the CPUC-ED, develop logic models and program theories to inform the evaluation plan.
3	The contractor works with the CPUC-ED on the development of the draft evaluation plan (with possible input from the program implementer) consistent with the ETP Protocol. As necessary, the plan must comply with the other Protocols (Impact Evaluation Protocol, Process Evaluation Protocol, Market Effects Protocols, the Sampling and Uncertainty Protocol and the Reporting Protocol) in the development of the evaluation plan and in the implementation and reporting efforts.
4	The CPUC-ED works with the evaluation contractor to finalize and approve an evaluation plan from which the contractor can begin the evaluation effort.
5	The contractor carries out all eight of the required Protocol requirements in order to measure key short, intermediate, and long-range performance indicators identified in the logic model.
6	The contractor reports the results of the final evaluation to the CPUC-ED and Joint Staff consistent with the provisions in the Reporting Protocol.
7	Once the report is accepted by the CPUC-ED, the contractor develops abstracts and posts the report on CALMAC web site following the CALMAC posting instructions.

Codes and Standards

Summary of Protocol-Driven Codes and Standards Evaluation Activities	
1	Joint staff selects an evaluation contractor to implement the Codes and Standards Program evaluation.
2	The evaluation contractor reviews the program change theories and the program logic models, identifies the technologies or behaviors that can be evaluated via the Protocol, constructs a draft evaluation plan and submits the plan for approval to the CPUC-ED. The contractor works with the CPUC-ED on the development of the draft evaluation plan and rigor levels. The plan must use the Impact Evaluation Protocol, the Sampling and Uncertainty Protocol and the Reporting Protocol in the development of the evaluation plan and in the implementation and reporting efforts.
3	The CPUC-ED works with the evaluation contractor to finalize and approve an evaluation plan from which the contractor can begin the evaluation effort.
4	The contractor conducts an assessment of the gross market-level energy impacts for each code and standard covered technology or behavior being evaluated consistent with the rigor level assignments.
5	The contractor determines the influence of the program on the adoption of each code and standard covered in the study and allocates adoption attribution. The assessment uses an interview approach for this assessment. This assessment is accomplished as early in the code change cycle as possible but preferably in the technology selection and demonstration phase of the cycle.
6	The contractor estimates naturally occurring code and standard covered technology or behavior adoption rates based on literature reviews and interviews with experts.
7	The contractor adjusts the gross market level energy savings estimates to account for the net adjustment factors for naturally occurring technology adoption, naturally occurring code change, and non-compliance. This approach nets out the influence of non-program-induced impacts from the gross market-level impacts for each technology.
8	The contractor estimates the timeline associated with adoption of a code and standard without the program, using a Delphi approach with an expert panel.
9	The program administrators remove savings estimates from their programs for code-covered measures.
10	The evaluation contractor assesses the construction and sales efforts for each utility company service territory and allocates savings by IOU based on the construction and sales estimates.
11	The contractor reports the results of the evaluation to the CPUC-ED and Joint Staff consistent with the provisions in the Reporting Protocol.
12	Once the report is accepted by the CPUC-ED, the contractor develops abstracts and posts the report on the CALMAC web site following the CALMAC posting instructions.

	posts the report on the CALMAC web site following the CALMAC posting instructions.
13	As needed, the CPUC-ED or the Joint Staff can request the evaluation contractor to update and report the actual energy savings over time consistent with the Protocol. Updates can be conducted with a different evaluation contractor than those doing the original assessment.

Effective Useful Life

Required Protocols for Measure Retention Study	
Rigor Level	Retention Evaluation Allowable Methods
Basic	1. In-place and operable status assessment based upon on-site inspections. Sampling must meet the Basic Rigor Level requirements discussed in this Protocol and must meet the requirements of the Sampling and Uncertainty Protocol. (The sampling requirements of this Protocol may need to meet the sampling requirements for the subsequent EUL study. See below specification.) 2. Non-site methods (such as telephone surveys/interviews, analysis of consumption data, or use of other data, e.g. from EMS systems) may be proposed but must be explicitly approved by Joint Staff through the evaluation planning process. Sampling must meet the Basic Rigor Level requirements discussed in this Protocol and must meet the requirements of the Sampling and Uncertainty Protocol. (The sampling requirements of this Protocol may need to meet the sampling requirements for the subsequent EUL study. See below specification.)
Enhanced	1. In-place and operable status assessment based upon on-site inspections. Sampling must meet the Enhanced Rigor Level requirements discussed in this Protocol and must meet the requirements of the Sampling and Uncertainty Protocol. (The sampling requirements of this Protocol may need to meet the sampling requirement for the subsequent EUL study. See below specification.)

Required Protocols for Degradation Study	
Rigor Level	Allowable Methods for Degradation Studies
Basic	1. Literature review required for technical degradation studies across a range of engineering-based literature, to include but not limited to manufacturer's studies, ASHRAE studies, and laboratory studies. Review of technology assessments. Assessments using simple engineering models for technology components and which examine key input variables and uncertainty factors affecting technical degradation. 2. Telephone surveys/interviews with a research design that meets accepted social science behavioral research expectations for behavioral degradation.
Enhanced	1. For technical degradation: field measurement testing. 2. For behavioral degradation: field observations and measurement.

Required Protocols for EUL Analysis Studies	
Rigor Level	Allowable Methods for EUL Analysis Studies
Basic	1. Classic survival analysis (defined below) or other analysis methods that specifically control for right-censored data (those cases of failure that might take place some time after data are collected) must be attempted. For methods not accounting for right-censored data, the functional form of the model used to estimate EUL ("model functional form") must be justified and theoretically supported. Sampling must meet the Basic Rigor Level requirements discussed in this Protocol and must meet the requirements of the Sampling and Uncertainty Protocol. Sample size requirements will be determined through the use of power analysis, results from prior studies on similar programs, and professional judgment. Power analysis used to determine the required sample size must be calculated by setting <u>power to at least at 0.7</u> to determine the sample size required at a 90% confidence level (alpha set at 0.10). Where other analyses or combined functional forms are used, power analysis should be set at these parameters to determine required sample sizes for regression-based approaches and a 90% confidence level with <u>30% precision</u> is to be used for non-regression components.
Enhanced	1. Classic survival analysis (defined below) or other analysis methods that specifically control for right-censored data (those cases of failure that might take place some time after data are collected) must be attempted. The functional form of the model used to estimate EUL ("model functional form") must be justified and theoretically supported. Sampling must meet the Enhanced Rigor Level requirements discussed in this Protocol and must meet the requirements of the Sampling and Uncertainty Protocol. Sample size requirements will be determined through the use of power analysis, results from prior studies on similar programs, and professional judgment. Power analysis used will set <u>power to at least to 0.8</u> to determine the sample size required at a 90% confidence level (alpha set at 0.10). Where other analyses or combined functional forms are used, power analysis should be set at these parameters to determine required sample sizes for regression-based approaches and a 90% confidence level with <u>10% precision</u> is to be used

	for non-regression components.

Summary of Protocol-Driven Impact Evaluation Activities	
1	Joint Staff will review retention, EUL, and degradation planning information, perhaps through an initial study of (1) prior retention, EUL, and degradation studies and methods, (2) required retention, EUL, and degradation sample sizes, and (3) assessment of data collection methods for the prioritized measure and delivery strategy/application needs. Along with any risk analysis information, Joint Staff will identify which measures by delivery strategy/application will receive which type of retention, EUL, and degradation evaluation, when, and at what rigor level. Joint Staff will determine any special needs on a case-by-case basis that will be required for particular retention, EUL, and degradation evaluations. Joint Staff will develop preliminary RFPs for groups of studies based upon timing of the needed data collection or analysis, similar sectors or issues to be addressed, and requiring similar skill sets. CPUC-ED will issue RFPs for retention, EUL, and degradation evaluations, select evaluation contractors, and establish scope(s) of work.
2	Evaluators will develop a research design and sampling plan to meet Protocol requirements as designated by the Joint Staff rigor level assignments. This includes meeting requirements from the Sampling and Uncertainty Protocol, as are applicable given Effective Useful Life Evaluation Protocol requirements. Research design and sampling must be designed to meet any of the Joint Staff requirements for additional analyses to include but not limited to areas designated of specific concern by the Joint Staff. Evaluators will develop and submit an Evaluation Plan to Joint Staff, and the plan will be revised as necessary to have an approved Evaluation Plan that meets the Effective Useful Life Evaluation Protocol.
3	All retention, EUL, and degradation study evaluation teams (including panel data collection teams) will make sure their teams are appropriately staffed, in order to meet the skills required for the research design, sampling, and selected retention, EUL, and degradation evaluation method, uncertainty analysis, and reporting being planned and conducted.
4	All retention, EUL, and degradation study evaluations will be planned, conducted, and analyzed to minimize potential bias in the estimates (showing the methods for doing this), and evaluators will report all analyses of potential bias issues as described in the Sampling and Uncertainty Protocol.
5	All retention, EUL, and degradation evaluations will be conducted according to the Evaluation Plan and appropriate Protocols.
6	Evaluators will develop the draft evaluation report in accordance to guidance provided by the Joint Staff and reporting requirements in this Protocol.
7	Final evaluation report will be developed in accordance to guidance provided by the Joint Staff, and then submitted to Joint Staff.
8	Once accepted by Joint Staff, abstracts will be developed, and a report will be posted on the CALMAC web site following the CALMAC posting instructions.

Summary of Protocol-Driven Market Effects Evaluation Activities

Required Protocols for Market Effects Evaluation Scoping Studies	
Level of Rigor	Scoping Study Requirements
Basic	Define the market by its location, the utilities involved, the equipment, behaviors, sector and the program years of interest. Develop market theory. Identify available secondary data and potential sources for primary data. Outline data collection and analysis approaches
Enhanced	Define the market by its location, the utilities involved, the equipment, behaviors, sector and the program years of interest. Develop market theory and logic model. Detail indicators. Identify available secondary data and primary data that can be used to track changes in indicators. Outline data collection approach. Recommend hypotheses to test in the market effects study. Recommend the analysis approach most likely to be effective.

Required Protocol for Market Theory and Logic Models	
Level of Rigor	Market Theory and Logic Model Requirements
Basic	Identification of assumptions about anticipated changes in the market and associated research questions. Market theory should include market operations and conditions, external influences, and assumptions about changes in the market (which could include market operational theory, market structure and function studies, and product and communication flows). Develop program theory and logic models across programs in that market. Analyze across both of these to examine program interventions, external influences and associated research questions. Theories and logic models should be generated through interviews with program staff and a sample of market actors.
Enhanced	Articulate market theory and, if reasonable, develop graphical model of market theory. Market theory should include market operations and conditions, and changes occurring in the market (could include market operational theory, market structure and function studies, and product and communication flows). Develop multiple program theory and logic models for those programs intervening in the market. Integrate the market theory and program theory/logic models to examine external and programmatic influences, assumptions about changes in the market and associated research questions. Theories and logic models should be generated through interviews or workshops with program staff from each of the programs and a sample of a wide variety of market actors. Use a literature review and other studies of these markets and iteration with program staff to ensure thoroughness in measuring the critical parameters for both market development from external influences and market effects.

Required Protocol for Market Effects Evaluation Indicator Studies	
Level of Rigor	Indicator Study Requirements
Basic	Select appropriate market actor group for each indicator, survey representative samples of market actors able to report on each indicator from market experience. A baseline study must be conducted as early as possible. On-going tracking provides the basis for comparisons.
Enhanced	Select appropriate market actor group for each indicator. Conduct longitudinal study of representative samples of market actors able to report on each indicator from market experience. Samples weighted to represent known parameters in the population of interest. A baseline study must be conducted as early as possible, on-going tracking provides the basis for comparisons.

Required Protocol for Preponderance of Evidence Approach to Causal Attribution Estimation	
Level of Rigor	Preponderance of Evidence Approach Requirements
Basic	A representative sample of market actors surveyed or interviewed to provide self-reports on perceived changes in the market, attribution and the sustainability of those changes.
Enhanced	Quasi-experimental or experimental design with comparison groups using a representative sample of market actors surveyed or interviewed to provide self-reports on perceived changes in the market, attribution and the sustainability of those changes.

Summary of Protocol-Driven Market Effects Evaluation Activities	
1	Joint staff identifies the markets or market sectors (and the associated set of programs) that will receive a market effects evaluation and identifies the potential approach and rigor level for the scoping study.
2	Joint staff identifies market- or market sector-specific study needs that will be assessed (including program-specific or program group specific study needs) from the evaluation. CPUC-ED issues request for proposals for market effects scoping study, selects the scoping study contractor and establishes a scope(s) of work.
3	Evaluation contractor develops scoping study. A scoping study will more finely define the market boundaries for the study, including its location, the utilities involved, the equipment or behaviors to be assessed and the program-influenced years of interest. The scoping study will develop a market theory and a logic model; identify the market change indicators to track; and the available primary and secondary data sources. The study will also identify the hypotheses to test and the data collection approach, and provide a recommended analysis approach and model specification (if appropriate).
4	A market change theory and logic model (MCT/LM) should be developed to identify assumed direction of effects and indicators for measuring effects. The market theory should include market operations and conditions, and changes occurring in the market (could include a market operations theory, market structure and function scenarios, and product and communication flows) The theory and logic model should be generated through interviews or workshops with program staff from each of the programs that are expected to influence the market being

	assessed and a sample of a wide variety of market actors and should incorporate a literature review.
5	Joint staff reviews the scoping study and determines how to proceed with the Market Effects Evaluation. CPUC-ED issues request for proposals for evaluation contractors, selects the contractor, establishes a final scope(s) of work and negotiates the contract.
6	All market effects evaluation teams must be staffed to meet the skills required for the research design, sampling, appropriate and selected evaluation method, uncertainty analysis and reporting requirements.
7	A research design and sampling plan should be developed to meet Protocol requirements at the market level to meet the Joint Staff assigned study rigor level. This includes meeting requirements from the Sampling and Uncertainty Protocol and the Reporting Protocol, as applicable. The evaluation contractor will develop an Evaluation Plan, submit it to the CPUC-ED and revise as necessary.
8	Indicators studies conducted as part of the Market Effects Evaluation should be based on the results of the scoping study, address the appropriate market actor group(s) for each indicator.
9	All Market Effects Evaluations must meet the requirements of the Sampling and Uncertainty Protocol. The 90/10 level of precision is a minimum precision target for the most important data collection efforts on its most important variables. Which data collection efforts and variables are considered to be the most important will be determined in close collaboration with the CPUC-ED
10	The gross market effects and the estimate of energy savings associated with the market effects should be estimated. Estimation of gross market effects can be as simple as comparing indicators between time one and time two and then multiplying the energy value derived in an M&V supported impact assessment or from DEER, or using a CPUC-ED-approved net energy effects model.
11	Attribution or causality should be addressed to estimate net effects using either a preponderance of evidence approach or a net effects modeling approach. c. For a preponderance of evidence approach a determination of attribution should use quasi-experimental or experimental design with comparison groups using a representative sample of market actors. This may include interviews to provide self-reports on perceived changes in the market, attribution and the sustainability of those changes as well as direct observation or other data to support changes resulting from the program. d. For a net effects modeling approach to estimate causality, the model specifications must reflect the complexity of the market. It is likely that such an approach will require multiple equations to model the various activities that occur in a market and the various points of intervention that energy efficiency programs exert on a market.
12	Sustainability should be addressed using a preponderance of evidence approach.
13	Develop draft evaluation report to include meeting all requirements in the Reporting Protocol and incorporating the program's performance metrics.
14	Develop final evaluation report in accordance to guidance provided by Joint Staff.

15	Submit final evaluation report to the CPUC-ED.
16	Once the report is accepted by the CPUC-ED, develop abstracts and post them and the report on CALMAC Web site following the CALMAC posting instructions

Sampling and Uncertainty

Required Protocols for Gross Impacts	
Rigor Level	Gross Impact Options
Basic	Simplified Engineering Models: The relative precision is 90/30 ¹⁵¹ . The sampling unit is the premise. The sample size selected must be justified in the evaluation plan and approved as part of the evaluation planning process.
	Normalized Annual Consumption (NAC) Models: There are no targets for relative precision. This is due to the fact that NAC models are typically estimated for all participants with an adequate amount of pre- and post-billing data. Thus, there is no sampling error. However, if sampling is conducted, either a power analysis ¹⁵² or justification based upon prior evaluations of similar programs must be used to determine sample sizes. The sample size selected must be justified in the evaluation plan and approved as part of the evaluation planning process.
Enhanced	Regression: There are no relative precision targets for regression models that estimate gross energy or demand impacts. Evaluators are expected to conduct, at a minimum, a statistical power analysis as a way of initially estimating the required sample size. ¹⁵³ Other information can be taken into account such as professional judgment and prior evaluations of similar programs. The sample size selected must be justified in the evaluation plan and approved as part of the evaluation planning process.
	Engineering Models: The target relative precision for gross energy and demand impacts is 90/10. The sampling unit is the premise. The sample size selected must be justified in the evaluation plan and approved as part of the evaluation planning process.

¹⁵¹ Also of interest, in addition to the relative precision, are the actual kWh, kW, and therm bounds of the interval.

¹⁵² Statistical power is the probability that statistical significance will be attained, given that there really is a treatment effect. Power analysis is a statistical technique that can be used (among other things) to determine sample size requirements to ensure statistical significance can be found. Power analysis is only being required in the Protocol for determining required sample sizes. There are several software packages and calculation Web sites that conduct the power analysis calculation. One of many possible references includes: Cohen, Jacob (1989) *Statistical Power Analysis for the Behavioral Sciences*, Lawrence Erlbaum Associates, Inc.

¹⁵³ Ibid.

Required Protocols for Net Impacts	
Rigor Level	Net Impacts Options
Basic	For the self-report approach (Option Basic.1), given the greater issues with construct validity and variety of layered measurements involved in estimating participant NTGRs, no relative precision target has been established. ¹⁵⁴ To ensure consistency and comparability a minimum sample size of 300 sites (or decision-makers in cases where decision-makers cover multiple sites) or a census ¹⁵⁵ , whichever is smaller, is required.
Standard	If the method used for estimating net energy and demand impacts is regression-based, there are no relative precision targets. If the method used for estimating NTGRs is regression-based (discrete choice), there are no relative precision targets. In either case, evaluators are expected to conduct, at a minimum, a statistical power analysis as a way of initially estimating the required sample size.¹⁵⁶ Other information can be taken into account such as professional judgment and prior evaluations of similar programs. For the self-report approach (Option Standard.2), there are no precision targets since the estimated NTGR will typically be estimated using information collected from multiple decision-makers involving a mix of quantitative and qualitative information around which a standard error cannot be constructed. Thus to ensure consistency and comparability, for such studies, a minimum sample size of 300 sites (or decision-makers in cases where decision-makers cover multiple sites) or a census, whichever is smaller, is required.
Enhanced	The requirements described for Enhanced apply depending on the methods chosen.

¹⁵⁴ This is considered the best feasible approach at the time of the creation of this Protocol. Like the other approaches to estimating the net-to-gross ratio (NTGR), there is no precision target when using the self-report method. However, unlike the estimation of the required sample sizes when using the regression and discrete choice approaches, the self-report approach poses a unique set of challenges to estimating required sample sizes. These challenges stem from the fact that the self-report methods for estimating free-ridership involve greater issues with construct validity, and often include a variety of layered measurements involving the collection of both qualitative and quantitative data from various actors involved in the decision to install the efficient equipment. Such a situation makes it difficult to arrive at a prior estimate of the expected variance needed to estimate the sample size.

Alternative proposals and the support and justifications that address all of the issues discussed here on the aggregation of variance for the proposed self-report method may be submitted to Joint Staff as an additional option (but not instead of the Protocol requirements) in impact evaluation RFPs and in Evaluation Plans. Joint Staff may elect to approve an Evaluation Plan with a well-justified alternative.

¹⁵⁵ A census is rarely achieved. Rather, one *attempts* to conduct a census, recognizing that there will nearly always be some sites, participants or non-participants who drop out for a variety of reasons such as refusals or insufficient data.

¹⁵⁶ Statistical power is the probability that statistical significance will be attained, given that there really is a treatment effect. Power analysis is a statistical technique that can be used (among other things) to determine sample size requirements to ensure statistical significance can be found. Power analysis is only being required in the Protocol for determining required sample sizes. There are several software packages and calculation Web sites that conduct the power analysis calculation.

Required Protocols for Measure-level Measurement and Verification	
Rigor Level	M&V Options
Basic	Simplified Engineering Models: The target relative precision for gross energy and demand impacts is 90/30. The sample unit may be the individual measure, a particular circuit or point of control as designated by the M&V plan.
Enhanced	Direct Measurement and Energy Simulation Models: The target relative precision for gross energy and demand impacts is 90/10. The sample unit may be the individual measure, a particular circuit or point of control as designated by the M&V plan.

Required Protocols for Sampling of Measures Within a Site
The target relative precision is 90/20 for each measure selected for investigation. The sampling unit (measure, circuit, control point) shall be designated by the M&V plan. The initial assumption regarding the coefficient of variation for determining sample size is 0.5.

Required Protocols for Verification	
Rigor Level	Verification Options
Basic	The target relative precision is 90/10. The key parameter upon which the variability for the sample size calculation is based is binary (i.e., Is it meeting the basic verification criteria specified in the M&V Protocol?).
Enhanced	The target relative precision is 90/10. The key parameter upon which the variability for the sample size calculation is based is binary (i.e., Is it meeting the enhanced verification criteria specified in the M&V Protocol?).

EVO

EFFICIENCY VALUATION ORGANIZATION



International Performance Measurement and Verification Protocol

Concepts and Options for Determining
Energy and Water Savings
Volume 1

Prepared by Efficiency Valuation Organization
www.evo-world.org

January 2012

EVO 10000 – 1:2012

International Performance Measurement and Verification Protocol

Concepts and Options for Determining
Energy and Water Savings
Volume 1

Prepared by Efficiency Valuation Organization
www.evo-world.org

January 2012

EVO 10000 – 1:2012



EVO Vision

A global marketplace that correctly values the efficient use of natural resources and utilizes end-use efficiency options as a viable alternative to supply options

EVO Mission

To develop and promote the use of standardized protocols, methods and tools to quantify and manage the performance risks and benefits associated with end-use energy-efficiency, renewable-energy, and water-efficiency business transactions



January 2012

Dear Readers,

As the world is coming to recognize that energy efficiency is foundational to good environmental management, the importance of proper savings documentation has never been greater. It is certainly in everyone's interest that predicted savings are achieved and properly reported.

Notably:

- **energy users** need to have robust methods of verifying achievement of their energy policy objectives, to get or maintain ISO 50001 certification for their management practices;
- **potential purchasers of energy efficiency products or services** want to know that their potential purchases have already proven themselves using widely recognized methods;
- **actual purchasers of energy efficiency products or services** need feedback on the effectiveness of their purchases, to help them fine tune performance and to inform further purchases;
- **governments and utilities** need to know that savings reported from energy efficiency programs are grounded in actual field-measured results following a widely accepted protocol.

Basically, the knowledge that energy savings can be transparently reported is vital to the acceptance of energy efficiency proposals.

EVO is the only organization dedicated to provision of tools for this purpose. This IPMVP, now in its seventh edition, **defines transparency in savings reports**, while assembling best practice from around the world. EVO also publishes the IEEFP (International Energy Efficiency Financing Protocol) to help energy efficiency investors identify and invest in well managed energy cost saving projects.

IPMVP's flexible framework of M&V Options allows practitioners to craft the right M&V Plan for their building or industrial facility, inspiring confidence in those who wish to harvest their financial and/or environmental benefits. Clear definition of terms, and heavy emphasis on consistent and transparent methods are the core precepts of the IPMVP. Though application details are unique to each project, IPMVP's flexible framework has been successfully applied to all types of energy efficiency techniques, for thousands of projects and programs, large and small, around the world.

IPMVP is the work of numerous volunteers and sponsors, listed herein and in previous editions. I would like to thank all those shown in the Acknowledgments section herein. You can join this truly unique group of professionals by submitting comments, joining an EVO committee, or supporting the movement by subscribing to EVO. I encourage all readers to provide feedback, so we can continuously improve the IPMVP (email to: ipmvpdev@evo-world.org).

John Cowan
Chair of the Board
Toronto, Canada

TABLE OF CONTENTS

Table of Contents	i
Acknowledgements	v
Changes In This Edition	vii
Preface.....	viii
Outline Of This Document	viii
Efficiency Valuation Organization and IPMVP	ix
EVO's Current Publications.....	ix
History Of Previous Editions	x
Training And Certification	x
EVO's Future Plans	x
Chapter 1 Introduction To IPMVP	1
1.1 Purpose And Scope Of IPMVP.....	1
1.2 Benefits Of Using IPMVP	1
1.3 IPMVP's Relationship To Other <i>M&V</i> Guidelines	2
1.4 Who Uses IPMVP?.....	2
Chapter 2 Definition and Purposes of <i>M&V</i>.....	4
2.1 Purposes Of <i>M&V</i>	5
Chapter 3 Principles of <i>M&V</i>	6
Chapter 4 IPMVP Framework And Options	7
4.1 Introduction	7
4.2 Energy, Water and Demand Terminology	8
4.3 The <i>M&V</i> Design and Reporting Process	8
4.4 Operational Verification	9
4.5 Savings Verification	10
4.6 Overview Of IPMVP Options	16

4.7 Options A & B: Retrofit Isolation	18
4.8 Option C: Whole Facility.....	25
4.9 Option D: Calibrated Simulation	28
4.10 Option Selection Guide	32
4.11 Persistence of Savings.....	35
Chapter 5 M&V Plan Contents	36
Chapter 6 M&V Reporting.....	39
Chapter 7 Adherence with IPMVP	40
Chapter 8 Other Common M&V Issues.....	41
8.1 Applying Energy Prices.....	41
8.2 Baseline Adjustments (<i>Non-Routine</i>).....	42
8.3 The Role Of Uncertainty (Accuracy).....	43
8.4 Cost	44
8.5 Balancing Uncertainty And Cost	45
8.6 Verification by an Independent Verifier.....	47
8.7 Data for Emission Trading	47
8.8 Minimum Operating Conditions.....	48
8.9 Weather Data.....	48
8.10 Minimum Energy Standards.....	48
8.11 Measurement Issues.....	49
8.12 Significant Digits.....	52
Chapter 9 Definitions	55
Chapter 10 References.....	59
10.1 Resource Organizations.....	63
10.2 Measurement References.....	64
10.3 Calibration References	66
10.4 European and International Standards Supporting Energy Efficiency In Buildings	66
Appendix A Examples.....	68

A-1 Introduction	68
A-2 Pump/Motor Efficiency Improvement – Option A.....	68
A-3 Lighting Efficiency – Option A.....	70
A-4 Compressed-Air Leakage Management – Option B	75
A-5 Turbine/Generator Set Improvement – Option B	76
A-6 Boiler Efficiency Improvement – Option A.....	78
A-7 Multiple <i>ECM</i> With Metered <i>Baseline</i> Data – Option C	79
A-8 Multiple <i>ECMs</i> In A Building Without Energy Meters In The <i>Baseline Period</i> – Option D	82
A-9 New Building Designed Better Than Code – Option D.....	84

Appendix B Uncertainty 87

B-1 Introduction.....	87
B-2 Modeling.....	92
B-3 Sampling	97
B-4 Metering	99
B-5 Combining Components of Uncertainty	100
B-6 Example Uncertainty Analysis.....	104

Appendix C Region-Specific Materials 107

C-1 United States of America - April 2007, updated October 2011	107
C-2 France - Juillet 2009	109
C-3 España - 2009.....	110
C-3.1 Catalunya - Setembre 2010	111
C-4 Romania - July 2010	112
C-5 Bulgaria - July 2010.....	113
C-6 Czech Republic - September 2010.....	114
C-7 Croatia - September 2010.....	115
C-8 Poland - September 2010	116

Appendix D USERS OF IPMVP 117

D-1 Energy-Performance Contractors and Their Building Customers	117
D-2 Energy-Performance Contractors and Their Industrial Customers	118
D-3 Industrial and Building Energy Users Doing Their Own Retrofits.....	119
D-4 Facility Managers Accounting For Energy/Water Budget Variances	119
D-5 New Building Designers.....	120

D-6	New Building Designers Seeking Recognition for Their Sustainable Designs	120
D-7	Existing Building Managers Seeking Recognition for the Environmental Quality of their Building Operations	120
D-8	Regional Efficiency-Program Designers and Managers.....	120
D-9	Water-Efficiency Project Developers	121
D-10	Emission Trading Programs.....	121
D-11	Energy Users Seeking ISO 50001 Certification	121
INDEX		122

ACKNOWLEDGEMENTS

The IPMVP is primarily maintained by the following volunteers. EVO appreciates their work in receiving comments and developing changes for the 2011 edition. EVO appreciates the support and commitments made by the employers of all EVO volunteers.

EVO Board of Directors (2012)

John Cowan, Chair (Canada), Environmental Interface Limited
Thomas Dreessen, Vice Chair (USA), EPS Capital
John Stephen Kromer, Treasurer and Immediate Past Chair (USA), Independent Consultant
Anees Iqbal, Secretary (UK), Maicon Associates Ltd.
Timothy Hui (China), EcoTech International (ETI)
Jin Ruidong (China), Chiana National Center of Housing and Settlement Department
Patrick Jullian (France), GIMELEC
Satish Kumar (India), Schneider-Electric
Tienan Li (China), Center for Industrial Energy Efficiency (CIEE)
Pierre Langlois (Canada), Econoler
Steve Schiller (USA), Schiller Consulting

IPMVP Committee (2012)

David Jump, Chair (USA,) Quantum Energy Services & Technologies, Inc. (QuEST)
Ellen Franconi, Vice-Chair (USA), Rocky Mountain Institute
Thomas Adams (USA), AFCESA
Raja Chirumamilla (USA), Sain Engineering Associates, Inc.
LJ Grobler (South Africa), Energy Cybernetics Pty Ltd.
Sami Khawaja (USA), Cadmus Group Inc.
David Korn (USA), Cadmus Group Inc.
Ken Lau (Canada), BC Hydro
Daniel Magnet (Switzerland), IBTECH
Fernando Milanez (Brazil), INEE
Tracy Phillips (USA), 7th Gen Energy Solutions
Phil Voss (USA), National Renewable Energy Lab
Kevin Warren (USA), Warren Energy Engineering
Lia Webster (USA), Portland Energy Conservation Inc. (PECI)
Raj Khilnani (India), Freelance Energy Consultant
Shankar Earni (USA), Lawrence Berkeley National Laboratory
Vilnis Vesma (UK), Degrees Days Direct Limited
Max Zhang (China), SGS-CSTC Standards Technical Services Co., Ltd.

M&V Plan Sub-Committee (2012)

Shankar Earni, Chair (USA), Lawrence Berkeley National Laboratory
Ellen Franconi, Vice-Chair (USA), Rocky Mountain Institute
Salil Gogte (USA), Nexant Inc.
David Korn (USA), Cadmus Group Inc.

IPMVP Fellows

Jeff Haberl (USA) Texas Agricultural and Mechanical University
John Stephen Kromer (USA), Independent Consultant
Steven R. Schiller (USA) Schiller Consulting Inc.

EVO is also grateful to its many individual subscribers around the world and its current organizational subscribers (also found at www.evo-world.org):

ADENE – Agencia para a Energia (Portuguese Energy Agency)
BC Hydro
Bonneville Power Administration
Dana Energy Services Company (DESCO)
EDF Electricite de France
Energy Decisions Pty Ltd.
EEVS – Energy Efficiency Verification Specialists
EU.BAC – European Building Automation and Controls Association
Gas Natural Fenosa
HEP-ESCO d.o.o.
ITE – Instituto de Tecnología EléctricaNexant, Inc.
North American Energy Standards Board
Navigant Consulting Inc.
Ontario Power Authority
Quantum Energy Services & Technologies, Inc.
Sabien Technology Ltd.
San Diego Gas & Electric Company
Schneider Electric
Services Industriels de Genève
SGS-CSTC Standards Technical Services Co., Ltd, China
Southern California Edison
Taiwan Green Productivity Foundation
Université de Genève
Vanguards Power (Hong Kong) Limited

CHANGES IN THIS EDITION

The 2012 edition makes the following changes to the 2010 edition:

1. Added content to Appendix C-1 sections for USA. Moved all references to ASHRAE to Annex C.
2. Moved detailed descriptions of users of IPMVP (Chapter 1.4) to a new Appendix D.
3. Formally modified requirements to include “*operational verification*” in addition to “savings verification” in Chapters 4 and 5. Consolidated former Chapters 4.4, 4.5 and 4.6 in the new Savings Verification Chapter 4.5, and renumbered subsequent sections of Chapter 4.
4. Included concept of “owner’s project requirements for M&V” in Chapter 4.
5. Definition of baseline for new construction added.
6. The role of Equations 1f and 1g) were clarified in Chapter 4.9.3.
7. Concept of “Monitoring and Targeting,” and its relationship to M&V added. Its reference is made in a new Chapter 4.11 on Persistence of Savings.
8. Definitions of statistical and uncertainty concepts clarified in Appendix B.
9. Made minor corrections to typographical or wording errors and updated references made to this edition of IPMVP.
10. A Catalonian section was added to Appendix C-3 for Spain

PREFACE

Outline Of This Document

The International Performance Measurement and Verification Protocol (IPMVP) Volume I is a guidance document describing common practice in measuring, computing and reporting savings achieved by energy or water efficiency projects at end user facilities. The IPMVP presents a framework and four measurement and verification (M&V) Options for transparently, reliably and consistently reporting a project's saving. M&V activities include site surveys, metering of energy or water flow(s), monitoring of independent variable(s), calculation, and reporting. When adhering to IPMVP's recommendations, these M&V activities can produce verifiable savings reports.

The IPMVP is intended to be used by professionals as a basis for preparing savings reports. Each user must establish its own specific M&V Plan that addresses the unique characteristics of the project. The IPMVP is not a standard and thus there is no formal compliance mechanism for this document. Adherence with the IPMVP requires preparation of a project specific M&V Plan that is consistent with IPMVP terminology. It must name the IPMVP Option(s) to be used, metering monitoring and analysis methods to be used, quality assurance procedures to be followed, and person(s) responsible for the M&V.

IPMVP Volume I Chapters are organized as follows:

1. Introduces IPMVP and EVO.
2. Defines M&V, and lists eight uses for M&V techniques.
3. Lays the foundation of M&V by defining the underlying Principles of good M&V. The balance of the document summarizes common industry methods for implementing these fundamental Principles.
4. Defines the IPMVP Framework and its four Options. It presents the basic methodologies and adjustments to energy or water measurements needed to properly report savings. Tables 2 and 4, and Figure 4 summarize the Options and offer guidance in choosing amongst them for each application.
5. Lists the topics that should be contained in an M&V Plan and gives guidance on design decisions needed to make the M&V activity cost effective for all users of the savings reports.
6. Defines a means of specifying the use of IPMVP and of claiming adherence with it.
7. Presents key information that should be included in each savings report.
8. Lists many additional issues that commonly arise in M&V design or reporting.
9. Lists definitions of all italicized words in the document.
10. Provides a list of references and some useful other resources.

Appendix A provides 12 example applications of IPMVP, in varying levels of detail. It refers to EVO's website for detailed examples of M&V Plans and Savings Reports.

Appendix B summarizes basic uncertainty quantification techniques to guide decisions about the level of rigour suitable for each M&V process.

Appendix C contains region-specific materials for the United States of America, France, Spain (incl. Catalonia), Romania, Bulgaria, the Czech Republic, Croatia and Poland.

Appendix D is a Users Guide to help different types of readers understand common ways of applying the document

Efficiency Valuation Organization and IPMVP

This International Performance Measurement and Verification Protocol (IPMVP) is sponsored by the Efficiency Valuation Organization (EVO), a non-profit private corporation. EVO envisions a global marketplace that correctly values the efficient use of natural resources and utilizes end-use efficiency as a viable alternative to new energy supply. EVO's mission is to develop and promote standardized methods to quantify and manage the risks and benefits associated with business transactions on end-use energy efficiency, renewable energy, and water efficiency. EVO is a subscriber-based organization with supporters around the world.

EVO is grateful to its volunteers who develop and maintain EVO documents. Members of our current Board and Committees active in developing this document appear in the Acknowledgements section, above, along with Organizational Subscribers.

EVO maintains a website (www.evo-world.org) which contains:

- A subscribers' section with access to pre-release copies of some EVO documents, reference materials, newsletters, discussion forums, and links to other resources;
- Latest released editions of the documents, and archival editions;
- Lists of the current committee members and supporters;
- Invitation for comments on IPMVP documents to the email address of: ipmvp@evo-world.org ;
- Information on EVO's training and certification programs;

EVO documents should include unique methods from around the world. Therefore, EVO is developing international, regional groups to document international M&V methods. To participate as a volunteer or subscriber, please visit the EVO website, www.evo-world.org, for current contact information.

EVO's current activities and plans are summarized below.

EVO's Current Publications

Currently EVO has four publications available on its website:

IPMVP Volume I Concepts and Options for Determining Energy and Water Savings

Volume I defines terminology and suggests good practices for documenting the effectiveness of energy or water efficiency projects that are implemented in buildings and industrial facilities. These terms and practices help managers to prepare M&V Plans, which specify how savings will be measured for each project. The successful M&V Plan enables verification by requiring transparent reports of actual project performance.

IPMVP Volume II Indoor Environmental Quality (IEQ) Issues

Volume II reviews IEQ issues as they may be influenced by an energy efficiency project. It highlights good project design and implementation practices for maintaining acceptable indoor conditions under an energy efficiency project. It advises on means of measuring IEQ parameters to substantiate whether indoor conditions have changed from the conditions of the baseline when determining savings. Volume II has been archived, and remains available in the archives section of the public library on EVO's website.

IPMVP Volume III Applications

Volume III contains specific application guidance manuals for Volume I. The two current applications manuals address new building construction (Part I) and renewable energy additions

to existing facilities (Part II). This volume is expected to be an area of continued development as more specific applications are defined.

International Energy Efficiency Financing Protocol (IEEFP)

The IEEFP provides guidelines for local financing institutions around the world to evaluate and finance energy efficiency and savings-based renewable projects.

History Of Previous Editions

The first edition of IPMVP, entitled the North American Energy Measurement and Verification Protocol, was published in March 1996. It was modified in December 1997 then renamed the International Performance Measurement and Verification Protocol. Options A and B were changed substantially when IPMVP was re-published in 2001 and minor editorial changes were added in a 2002 edition. Volume II on Indoor Environmental Quality was published in 2002. Committees, sponsored by the United States' Department of Energy (DOE) wrote and edited these documents.

In 2002, IPMVP Inc. was incorporated as an independent non-profit corporation in order to include the international community and relieve the U.S. Department of Energy of its responsibilities as the organizer. IPMVP Inc. raised its own funds, created a website, and published the new Volume III Parts on New Construction and Renewables. In 2004, IPMVP Inc. was renamed Efficiency Valuation Organization as it expanded its focus.

In 2007, EVO updated IPMVP Volume I primarily for clarity, thought re-writing the uncertainty Appendix B. No substantive changes were made to the core concepts, though the titles of Options A & B were expanded along with their descriptive materials, to ensure proper understanding. In 2009, Volume I was modified to separate USA-specific references and to establish a structure for many region-specific materials in a new Appendix C for the USA and France. . In 2010, Volume I was modified to add more European references and Appendices, and updated the significant digitis presentation.

Training And Certification

EVO recognizes that documents alone will not improve the valuation of energy efficiency in the world. Therefore EVO and its world wide partners introduced awareness and training programs about measurement and verification. These programs educate professionals on methods and recent developments in M&V.

EVO also has a Certified Measurement and Verification Professional (CVMP) program for professionals who pass a test demonstrating their knowledge of IPMVP and have appropriate knowledge experience or training. CMVP[®]s should be competent to develop M&V Plans and to manage M&V programs for straightforward applications. For more information on the CMVP[®] program, and for the names of designated CMVP[®]s, visit www.evo-world.org .

EVO's Future Plans

EVO's subscribers, and volunteers determine its future plans to create new educational efforts and documents about efficiency valuation. EVO welcomes IPMVP readers to become EVO subscribers, provide recommendations, and participate in new and existing EVO activities.

In light of its international focus, EVO is in the process of:

- Developing active regional affiliates who contribute to the development and maintenance of EVO publications;
- Conducting additional training and certification programs around the world;
- Preparing its latest documents in a variety of languages; and

- Encouraging its internet-based community of subscribers to share efficiency-valuation ideas.

EVO welcomes feedback and suggestions. Please direct your comments by email to ipmvprev@evo-world.org. All comments will be considered, though EVO will not necessarily reply directly. The latest English version and certified translations of EVO documents will always be available for internet download at www.evo-world.org. EVO plans to revise this document every year. Please let us know how our services can be improved or expanded.

CHAPTER 1 INTRODUCTION TO IPMVP

1.1 Purpose And Scope Of IPMVP

Efficiency Valuation Organization (EVO) publishes the International Performance Measurement and Verification Protocol (IPMVP) to increase investment in energy and water efficiency, demand management and renewable energy projects around the world.

The IPMVP promotes efficiency investments by the following activities.

- IPMVP documents common terms and methods to evaluate performance of efficiency projects for buyers, sellers and financiers. Some of these terms and methods may be used in project agreements, though IPMVP does not offer contractual language.
- IPMVP provides methods, with different levels of cost and accuracy, for determining savings¹ either for the whole facility or for individual energy conservation measures (ECM)²;
- IPMVP specifies the contents of a Measurement and Verification Plan (M&V Plan). This M&V Plan adheres to widely accepted fundamental principles of M&V and should produce verifiable savings reports. An M&V Plan must be developed for each project by a qualified professional³.
- IPMVP applies to a wide variety of facilities including existing and new buildings and industrial processes. Chapter 1.4, User's Guide, summarizes how different readers might use IPMVP.

IPMVP Volume I defines *M&V* in Chapter 2, presents the fundamental principles of *M&V* in Chapter 3, and describes a framework for a detailed *M&V Plan* in Chapter 4. The details of an *M&V Plan* and *savings* report are listed in Chapters 5 and 6, respectively. The requirements for specifying use of IPMVP or claiming adherence with IPMVP are shown in Chapter 7. Volume I also contains a summary of common M&V design issues, Chapter 8, and lists other *M&V* resources. Twelve example projects are described in Appendix A and basic uncertainty analysis methods are summarized in Appendix B. Region-specific materials are in Appendix C. Specific guidance for different types of users is in Appendix D.

IPMVP Volume II provides a comprehensive approach to evaluating building indoor-environmental-quality issues that are related to ECM design, implementation and maintenance. Volume II suggests measurements of indoor conditions to identify changes from conditions of the baseline period.

IPMVP Volume III provides greater detail on M&V methods associated with new building construction, and with renewable energy systems added to existing facilities.

IPMVP Volumes I and III are a living suite of documents, with the latest modifications available on EVO's website (www.evo-world.org). Volume II is now found in the archives of the EVO website.

1.2 Benefits Of Using IPMVP

IPMVP's history since 1995 and its international use brings the following benefits to programs that adhere to IPMVP's guidance.

¹ Words in italics have the special meanings defined in Chapter 8.

² Although there is some debate over the differences between two terms — energy conservation measure (ECM) and energy efficiency measure (EEM) — the common ECM term is defined to include both conservation and efficiency actions. See Chapter 8.

³ www.evo-world.org contains the current list of Certified *M&V* Professionals (CMVP[®]s), persons with appropriate experience and who have demonstrated their knowledge of IPMVP by passing an examination.

- Substantiation of payments for performance. Where financial payments are based on demonstrated energy or water *savings*, adherence to IPMVP ensures that *savings* follow good practice. An IPMVP-adherent *savings report* allows a customer, an energy user or a utility, to readily accept reported performance. *Energy service companies (ESCOs)* whose invoices are supported by IPMVP-adherent *savings reports*, usually receive prompt payments.
- Lower transaction costs in an *energy performance contract*. Specification of IPMVP as the basis for designing a project's *M&V* can simplify the negotiations for an *energy performance contract*.
- International credibility for energy *savings reports*, thereby increasing the value to a buyer of the associated energy *savings*.
- Enhanced rating under programs to encourage or label sustainably designed and/ or operated facilities.
- Help national and industry organizations promote and achieve resource efficiency and environmental objectives. The IPMVP is widely adopted by national and regional government agencies and by industry organizations to help manage their programs and enhance the credibility of their reported results.

1.3 IPMVP's Relationship To Other *M&V* Guidelines

Chapter 9 lists other interesting resources for readers of IPMVP. Appendix C lists other guidelines, protocols, and documents from different regions of the world that are applications of IPMVP, or provide references to relevant codes, standards, and programs that reference IPMVP.

1.4 Who Uses IPMVP?

IPMVP presents common principles and terms that are widely accepted as basic to any good *M&V* process. It does not define the *M&V* activities for all applications. Each project must be individually designed to suit the needs of all readers of energy or water *savings reports*. This individual design is recorded in the project's *M&V Plan* and *savings* are reported as defined therein.

This document is written to progressively provide greater levels of definition of *M&V* practice as it progresses through the Chapters as summarized below.

- Chapter 2 defines *M&V* and describes eight different applications for *M&V* techniques.
- Chapter 3 present the six foundational principles of good *M&V* practice and the IPMVP. They are useful for guiding *M&V* design details where IPMVP is silent.
- Chapter 4 presents the general framework and *savings* computation equations needed to properly express *savings*. Table 2 summarizes four *M&V* design Options and Chapters 4.7 - 4.9 describe each of them. Chapter 4.10 offers guidance and a logic diagram for selecting the right Option for any application. Appendix A provides example applications of IPMVP's methods to 12 typical projects.
- Chapter 5 lists the topics and data which should be included in an *M&V Plan* and offers some suggestions on key issues which might be discussed under each topic. Readers can use this as a checklist for describing the *M&V* design for a particular project.
- Chapter 6 lists the topics and data that should be included in *savings* reports.
- Chapter 7 shows the requirements for claiming adherence with IPMVP and suggests terms for specifying the use of IPMVP in contracts.

- Chapter 8 reviews a variety of common *M&V* issues that need to be considered in any *M&V* program. A key issue governing the design and operation of an *M&V* system is the competing needs for reasonable accuracy and reasonable cost. Each user must find its own balance between the accuracy and cost of reporting. Chapter 8.5 particularly focuses on the factors involved in this tradeoff. Appendix B provides an overview of some uncertainty and statistical methods, but this overview is not a definitive text on the topic. Users are advised to seek appropriate statistical design help for any *M&V* program data normalization, sampling or uncertainty evaluation techniques they may use. Chapter 8 also presents design issues surrounding metering for *M&V* programs, though it is not a definitive text on metering.
- Chapter 9 contains the definitions of key terms used in this document. The terms are italicized throughout the document to indicate that they have the special meanings given in Chapter 9.
- Chapter 10 lists useful readings, references, and other sources of useful material.

Though the application of IPMVP is unique to each project, certain types of users will have similar methods in their *M&V Plans* and implementation. Appendix D.1 through D.10 point out some of the key ways this document may be used by the following user groups:

- Energy performance contractors and their building customers
- Energy performance contractors and their industrial process customers
- Energy users doing their own retrofits and wanting to account for *savings*
- Facility managers properly accounting for energy budget variances
- New building designers
- New building designers seeking recognition for the sustainability of their designs
- Existing building managers seeking recognition for the environmental quality of their building operations
- Utility demand side management program designers and managers
- Water efficiency project developers
- Emission reduction trading program designers
- Energy user's seeking ISO 50001 certification

Financial backers and purchasers of emission credits from any of the above applications will find the key ways to use this document under these same headings.

This Chapter uses terms explained in subsequent Chapters as noted in brackets, or as defined in Chapter 9 for italicized words.

CHAPTER 2 DEFINITION AND PURPOSES OF *M&V*

“*Measurement and Verification*” (*M&V*) is the process of using measurement to reliably determine actual *savings*⁴ created within an individual facility by an energy management program. *Savings* cannot be directly measured, since they represent the absence of *energy* use. Instead, *savings* are determined by comparing measured use before and after implementation of a project, making appropriate adjustments for changes in conditions.

M&V activities consist of some or all of the following:

- meter installation calibration and maintenance,
- data gathering and screening,
- development of a computation method and acceptable estimates,
- computations with measured data, and
- reporting, quality assurance, and third party verification of reports.

When there is little doubt about the outcome of a project, or no need to prove results to another party, applying *M&V* methods to calculate savings may not be necessary. However, it is still wise to verify (initially and repeatedly) that the installed equipment is able to produce the expected savings. Verification of the potential to achieve *savings* is referred to as *operational verification*, which may involve inspection, commissioning of equipment, functional performance testing and/or data trending (see Section 4.4. below). IPMVP-adherent *M&V* includes both *operational verification* and an accounting of *savings* based on site energy measurements before and after implementation of a project, and adjustments, as described above.

M&V is not just a collection of tasks conducted to help a project meet IPMVP requirements. Properly integrated, each *M&V* task serves to enhance and improve facility operation and maintenance of savings. As shown in Figure 1, *M&V* activities overlap with other project efforts (e.g. collecting data to both identify ECMs and establish energy baselines, commissioning and operational verification of installed ECMs, and installing monitoring systems to track and maintain savings persistence, etc.). Identifying these project synergies and establishing roles and responsibilities of involved parties during project planning will support a coordinated team effort. This can leverage complementary scopes and control *M&V*-related costs.

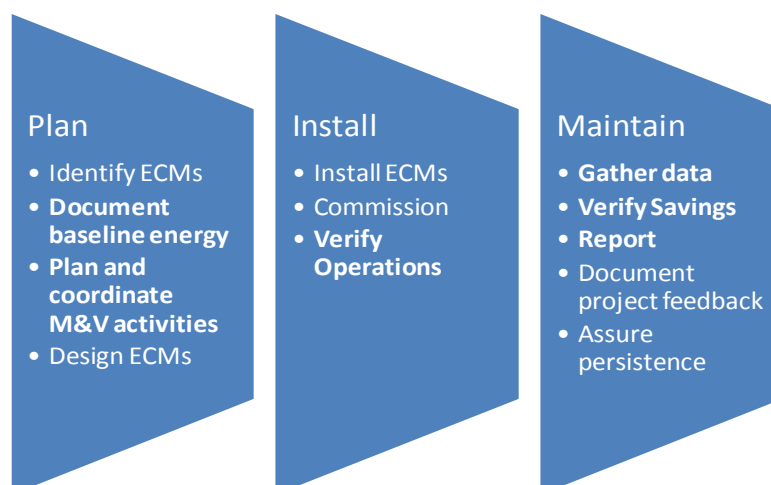


**Figure 1
Process
Timeline
(*M&V*
activities in
bold)**

⁴ Words in *italics* have the special meanings defined in Chapter 9.

2.1 Purposes Of M&V

M&V techniques can be used by facility owners or *energy* efficiency project investors for the following purposes:

a) Increase *energy savings*

Accurate determination of *energy savings* gives facility owners and managers valuable feedback on their *energy conservation measures (ECMs)*. This feedback helps them adjust *ECM* design or operations to improve *savings*, achieve greater persistence of *savings* over time, and lower variations in *savings* (Kats et al. 1997 and 1999, Haberl et al. 1996).

b) Document financial transactions

For some projects, the energy efficiency savings are the basis for performance-based financial payments and/or a guarantee in a performance contract. A well-defined and implemented *M&V Plan* can be the basis for documenting performance in a transparent manner and subjected to independent verification.

c) Enhance financing for efficiency projects

A good *M&V Plan* increases the transparency and credibility of reports on the outcome of efficiency investments. It also increases the credibility of projections for the outcome of efficiency investments. This credibility can increase the confidence that investors and sponsors have in *energy* efficiency projects, enhancing their chances of being financed.

d) Improve engineering design and facility operations and maintenance

The preparation of a good *M&V Plan* encourages comprehensive project design by including all *M&V* costs in the project's economics. Good *M&V* also helps managers discover and reduce maintenance and operating problems, so they can run facilities more effectively. Good *M&V* also provides feedback for future project designs.

e) Manage *energy* budgets

Even where *savings* are not planned, *M&V* techniques help managers evaluate and manage *energy* usage to account for variances from budgets. *M&V* techniques are used to adjust for changing facility-operating conditions in order to set proper budgets and account for budget variances.

f) Enhance the value of emission-reduction credits

Accounting for emission reductions provides additional value to efficiency projects. Use of an *M&V Plan* for determining *energy savings* improves emissions-reduction reports compared to reports with no *M&V Plan*.

g) Support evaluation of regional efficiency programs

Utility or government programs for managing the usage of an *energy* supply system can use *M&V* techniques to evaluate the savings at selected *energy* user facilities. Using statistical techniques and other assumptions, the *savings* determined by *M&V* activities at selected individual facilities can help predict savings at unmeasured sites in order to report the performance of the entire program.

h) Increase public understanding of energy management as a public policy tool

By improving the credibility of energy management projects, *M&V* increases public acceptance of the related emission reduction. Such public acceptance encourages investment in energy-efficiency projects or the emission credits they may create. By enhancing *savings*, good *M&V* practice highlights the public benefits provided by good *energy* management, such as improved community health, reduced environmental degradation, and increased employment

CHAPTER 3 PRINCIPLES OF *M&V*

The fundamental principles of good *M&V*⁵ practice are described below, in alphabetical order.

Accurate *M&V* reports should be as accurate as the *M&V* budget will allow. *M&V* costs should normally be small relative to the monetary value of the *savings* being evaluated. *M&V* expenditures should also be consistent with the financial implications of over- or under-reporting of a project's performance. Accuracy tradeoffs should be accompanied by increased conservativeness in any estimates and judgements.

Complete The reporting of *energy savings* should consider all effects of a project. *M&V* activities should use measurements to quantify the significant effects, while estimating all others.

Conservative Where judgements are made about uncertain quantities, *M&V* procedures should be designed to under-estimate *savings*.

Consistent The reporting of a project's *energy* effectiveness should be consistent between:

- different types of *energy* efficiency projects;
- different *energy* management professionals for any one project;
- different periods of time for the same project; and
- *energy* efficiency projects and new *energy* supply projects.

'Consistent' does not mean 'identical,' since it is recognized that any empirically derived report involves judgements which may not be made identically by all reporters. By identifying key areas of judgement, IPMVP helps to avoid inconsistencies arising from lack of consideration of important dimensions.

Relevant The determination of *savings* should measure the performance parameters of concern, or least well known, while other less critical or predictable parameters may be estimated.

Transparent All *M&V* activities should be clearly and fully disclosed. Full disclosure should include presentation of all of the elements defined in Chapters 5 and 6 for the contents of an *M&V Plan* and a *savings* report, respectively.

The balance of this document presents a flexible framework of basic procedures and four Options for achieving *M&V* processes which follow these fundamental principles. Where the framework is silent or inconsistent for any specific application, these *M&V* principles should be used for guidance.

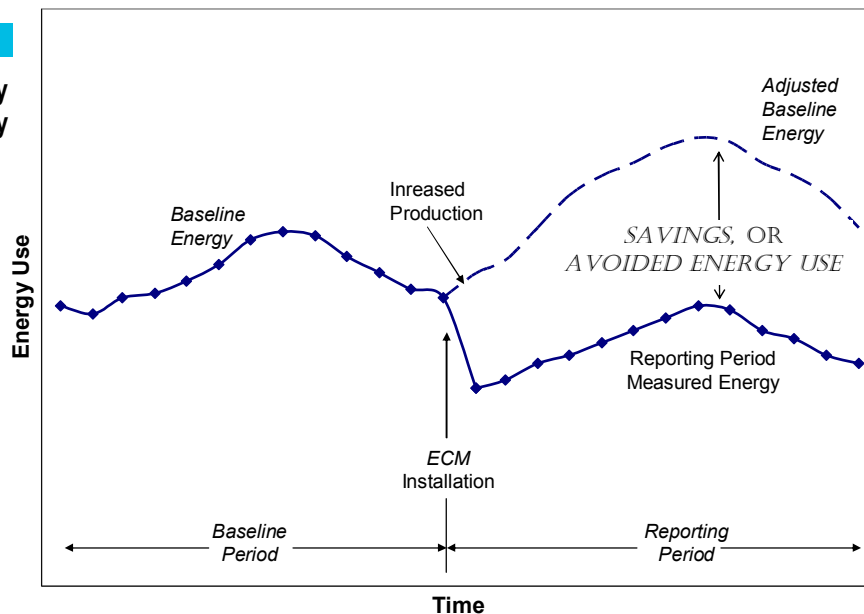
⁵ Words in italics have the special meanings defined in Chapter 9.

CHAPTER 4 IPMVP FRAMEWORK AND OPTIONS

4.1 Introduction

Energy, water or demand *savings*⁶ cannot be directly measured, since *savings* represent the absence of energy/water use or demand. Instead, *savings* are determined by comparing measured use or demand before and after implementation of a program, making suitable adjustments for changes in conditions.

Figure 2 Example Energy History



As an example of savings determination process, Figure 2 shows the energy-usage history of an industrial boiler before and after the addition of an *energy conservation measure (ECM)* to recover heat from its flue gases. At about the time of *ECM* installation, plant production also increased.

To properly document the impact of the *ECM*, its energy effect must be separated from the energy effect of the increased production. The “*baseline energy*” use pattern before *ECM* installation was studied to determine the relationship between energy use and production. Following *ECM* installation, this *baseline* relationship was used to estimate how much energy the plant would have used each month if there had been no *ECM* (called the “*adjusted-baseline energy*”). The *saving*, or ‘*avoided energy use*’ is the difference between the *adjusted-baseline energy* and the energy that was actually metered during the *reporting period*.

Without the adjustment for the change in production, the difference between *baseline energy* and *reporting period* energy would have been much lower, under-reporting the effect of the heat recovery.

It is necessary to segregate the *energy* effects of a *savings* program from the effects of other simultaneous changes affecting the *energy* using systems. The comparison of before and after energy use or demand should be made on a consistent basis, using the following general Equation 1):

⁶ Words in italics have the special meanings defined in Chapter 9.

$$\text{Savings} = (\text{Baseline-Period Use or Demand} - \text{Reporting-Period Use or Demand}) \pm \text{Adjustments}$$

1)

The "Adjustments" term in this general equation is used to re-state the use or demand of the *baseline* and *reporting periods* under a common set of conditions. This adjustments term distinguishes proper *savings* reports from a simple comparison of cost or usage before and after implementation of an *energy conservation measure (ECM)*. Simple comparisons of utility costs without such adjustments report only cost changes and fail to report the true performance of a project. To properly report "*savings*," adjustments must account for the differences in conditions between the *baseline* and *reporting periods*.

The baseline in an existing *facility* project is usually the performance of the *facility* or system prior to modification. This baseline physically exists and can be measured before changes are implemented. In new construction, the baseline is usually hypothetical and defined based on code, regulation, common practice or documented performance of similar *facilities*. In either case, the baseline model must be capable of accommodating changes in operating parameters and conditions so "adjustments" can be made.

The balance of this Chapter defines basic methods in these measurement and adjustment processes. If these methods do not cover all matters that arise in your project, consult the Principles of *M&V* (Chapter 3) for further guidance.

4.2 Energy, Water and Demand Terminology

The processes of determining *savings* in energy are analogous to those for determining *savings* in water or demand. To simplify the descriptions in this document, the italicized word *energy* will commonly be used to mean energy and water use or demand. Similarly the word *Energy Conservation Measure (ECM)* will commonly be used to mean measures to improve efficiency or conserve energy or water, or manage demand.

4.3 The *M&V* Design and Reporting Process

The *M&V* design and reporting process parallels the *ECM* design and implementation process. The *M&V* processes should involve the following steps:

1. Consider the needs of the user of the planned *M&V* report(s). If the user is focused on overall cost control, Whole-Facility methods may be most suited. If user focus is on particular *ECMs*, Retrofit Isolation techniques may be most suited (see Chapter 4.4).
2. While developing the *ECM(s)*, select the IPMVP Option (see Chapters 4.7 - 4.11) that best suits the *ECM(s)*, the needs for accuracy and the budget for *M&V*. Decide whether adjustment of all *energy* quantities will be made to the *reporting period* conditions or to some other set of conditions (see Chapter 4.6). Decide the duration of the *baseline period* and the *reporting period* (Chapter 4.5) (These fundamental decisions may be written into the terms of an *energy-performance contract*.)
3. Gather relevant *energy* and operating data from the *baseline period* and record them in a way that can be accessed in the future.
4. Prepare an *M&V Plan* (Chapter 5) containing the results of steps 1 through 3 above. It should define the subsequent steps 5 through 9.
5. As part of the final *ECM* design and installation, also design, install, calibrate and commission any special measurement equipment that is needed under the *M&V Plan*.

6. After the ECM is installed, ensure it has the potential to perform and achieve savings by conducting operational verification. This may include inspecting the installed equipment and revising operating procedures as needed to conform to the design intent of the *ECM*. This requirement may be fulfilled by a formal "commissioning" process as part of the project .
7. Gather energy and operating data from the *reporting period*, as defined in the *M&V Plan*.
8. Compute *savings* in *energy* and monetary units in accordance with the *M&V Plan*.
9. Report *savings* in accordance with the *M&V Plan* (see Chapter 6).

Interactive Effects - Example

For an ECM, which reduces the power requirements of electric lights, the *measurement boundary* should include the power to the lights. However lowering lighting energy may also lower any mechanical cooling requirements and/or raise any heating requirements. Such heating and cooling energy flows attributable to the lights cannot usually be easily measured. They are *interactive effects* which may have to be estimated, rather than included within the measurement boundary.

Steps 7 through 9 are repeated periodically when a *savings* report is needed.

A third party may verify that the *M&V Plan* adheres to IPMVP, and possibly a performance contract. This third party may also verify that *savings* reports comply with the approved *M&V Plan* (see Chapter 8.6).

4.4 Operational Verification

Operational Verification should be performed as part of any project M&V program. It serves as a low-cost initial step for realizing savings potential and should precede savings verification activities. A range of operational verification methods can be applied, as outlined in Table 2. Selection of a given approach depends on the ECM's characteristics as noted. However, it can also be influenced by the savings verification approach taken. For example, if Option B is being used to verify savings then a more simple visual inspection may suffice for operational verification. However if Option A is applied, then a more thorough *operational verification* approach should be used to verify that the ECM functionality is confirmed and characterization complete.

Operational verification activities are accomplished through comprehensive commissioning of affected systems supplemented by more data-driven activities (e.g. data trending and review). The M&V Plan should note the *operational verification* approach in addition to the savings verification method. *Operational verification* should be completed prior to implementing M&V savings verification activities. This ensures the savings from ECMs, energy-efficient controls and operation improvements are fully realized.

Table 1
Operational
Verification
Approaches

Operational Verification Approach	Typical ECM Application	Activities
Visual Inspection	ECM will perform as anticipated when properly installed; direct measurement of ECM performance is not possible. Examples: wall insulation, windows	View and verify the physical installation of the ECM
Sample Spot Measurements	Achieved ECM performance can vary from published data based on installation details or component load. Examples: fixtures/lamps/ballasts, fans, pumps	Measure single or multiple key energy-use parameters for a representative sample of the ECM installations
Short-Term Performance Testing	ECM performance may vary depending on actual load, controls, and/or interoperability of components. Examples: Daylighting sensors and lighting dimming controls, VSD fans, Demand-control ventilation	Test for functionality and proper control. Measure key energy-use parameters. May involve conducting tests designed to capture the component operating over its full range or performance data collection over sufficient period of time to characterize the full range of operation.
Data Trending and Control-Logic Review	ECM performance may vary depending on actual load and controls. Component or system is being monitored and controlled through the BAS or can be monitored through independent meters.	Set up trends and review data and/or control logic. Measurement period may last for a few days to a few weeks, depending on the period needed to capture the full range of performance.

Operational verification activities can also be applied following the *Reporting Period* to support energy savings persistence. While not formally a part of the M&V process, such a practice is beneficial for an organization which has improved its energy efficiency. It reduces the risk of adverse shifts in performance associated with ECMs that can fail, fade or be bypassed.

4.5 Savings Verification

The following sections add details about how to determine and report *savings*.

4.5.1 Measurement Boundary

Savings may be determined for an entire *facility* or simply for a portion of it, depending upon the purposes of the reporting.

- If the purpose of reporting is to help manage only the equipment affected by the *savings* program, a *measurement boundary* should be drawn around that equipment. Then all significant *energy* requirements of the equipment within the boundary can be determined⁷. This approach is used in the Retrofit Isolation Options of Chapter 4.7.
- If the purpose of reporting is to help manage total *facility energy* performance, the meters measuring the supply of *energy* to the total *facility* can be used to assess performance and *savings*. The *measurement boundary* in this case encompasses the whole *facility*. The Whole Facility Option C, is described in Chapter 4.8.
- If *baseline* or *reporting period* data are unreliable or unavailable, *energy* data from a calibrated simulation program can take the place of the missing data, for either part or all of the *facility*. The *measurement boundary* can be drawn accordingly. The Calibrated Simulation Option D is described in Chapter 4.9.

⁷ Determination of *energy* may be by direct measurement of energy flow or by direct measurement of *proxies* of *energy* use that give direct indication of *energy* use.

Some of the energy requirements of the systems or equipment being assessed may arise outside a practical *measurement boundary*. Nevertheless, all *energy effects* of the *ECM(s)* should be considered. Those energy effects that are significant should be determined from measurements, the rest being estimated or ignored.

Any energy effects occurring beyond the notional *measurement boundary* are called '*interactive effects*'⁸. Find a way to estimate the magnitude of these *interactive effects* in order to determine *savings*. Alternatively they may be ignored as long as the *M&V Plan* includes discussion of each effect and its likely magnitude.

4.5.2 Measurement Period Selection

Care should be taken in selecting the period of time to be used as the *baseline period* and the *reporting period*. Strategies for each are discussed below.

Baseline Period

The *baseline period* should be established to:

- Represent all operating modes of the *facility*. This period should span a full operating cycle from maximum energy use to minimum.

- Whole-building energy use can be significantly affected by weather conditions. Typically, a whole year of baseline data is needed to define a full operating cycle.
- The energy use of a compressed air system may only be governed by plant production levels, which vary on a weekly cycle. So one week's data may be all that is needed to define baseline performance.

- Fairly represent all operating conditions of a normal operating cycle. For example, though a year may be chosen as the baseline period, if data is missing during the selected year for one month, comparable data for the same month in a different year should be used to ensure the baseline record does not under represent operating conditions of the missing month.
- Include only time periods for which all fixed and variable energy-governing facts are known about the facility. Extension of baseline periods backwards in time to include multiple cycles of operation requires equal knowledge of all energy-governing factors throughout the longer baseline period in order to properly derive routine and non-routine adjustments (see Chapter 4.6) after ECM installation.
- Coincide with the period immediately before commitment to undertake the retrofit. Periods further back in time would not reflect the conditions existing before retrofit and may therefore not provide a proper *baseline* for measuring the effect of just the *ECM*.

ECM planning may require study of a longer time period than is chosen for the *baseline period*. Longer study periods assist the planner in understanding *facility* performance and determining what the normal *cycle* length actually is.

⁸ These *interactive effects* are sometimes called 'leakages.'

Reporting Period

The user of the *savings* reports should determine the length of the *reporting period*. The *reporting period* should encompass at least one normal operating *cycle* of the equipment or *facility*, in order to fully characterize the *savings* effectiveness in all normal operating modes.

Some projects may cease reporting *savings* after a defined "test" period ranging from an instantaneous reading to a year or two.

The length of any *reporting period* should be determined with due consideration of the life of the *ECM* and the likelihood of degradation of originally achieved *savings* over time.

Regardless of the length of the reporting period, metering may be left in place to provide feedback of operating data for routine management purposes and specifically to detect subsequent adverse changes in performance. Section 4.9 explains this at greater length.

If reducing the frequency of savings measurement after initial proof of performance, other on site monitoring activities could be intensified to ensure savings remain in place.

IPMVP-adherent *savings* can only be reported for the *reporting period* that uses IPMVP-adherent procedures. If IPMVP-adherent savings are used as a basis for assuming future *savings*, future *savings* reports do not adhere to IPMVP.

Adjacent Measurement Periods (On/Off Test)

When an *ECM* can be turned on and off easily, *baseline* and *reporting periods* may be selected that are adjacent to each other in time. A change in control logic is an example of an *ECM* that can often be readily removed and reinstated without affecting the facility.

Such "On/Off tests" involve *energy* measurements with the *ECM* in effect, and then immediately thereafter with the *ECM* turned off so that pre-*ECM* (*baseline*) conditions return. The difference in energy use between the two adjacent measurement periods is the *savings* created by the *ECM*. Equation 1) of Chapter 4.1 can be used for computing *savings*, without an adjustments term if all *energy*-influencing factors are the same in the two adjacent periods.

This technique can be applied under both Retrofit Isolation and Whole-Facility Options. However *measurement boundaries* must be located so that it is possible to readily detect a significant difference in metered *energy* use when equipment or systems are turned on and off.

The adjacent periods used for the On/Off test should be long enough to represent stable operation. The periods should also cover the range of normal facility operations. To cover the normal range, the On/Off test may need to be repeated under different operating modes such as various seasons or production rates.

Take notice that *ECMs* which can be turned Off for such testing also risk being accidentally or maliciously turned Off when intended to be On.

4.5.3 Basis For Adjustments

The adjustments term shown in Equation 1) of Chapter 4.1 should be computed from identifiable physical facts about the energy governing characteristics of equipment within the *measurement boundary*. Two types of adjustments are possible:

- **Routine Adjustments** – for any *energy*-governing factors, expected to change routinely during the *reporting period*, such as weather or production volume. A variety of techniques can be used to define the adjustment methodology. Techniques may be as simple as a constant value (no adjustment) or as complex as a several multiple parameter non-linear equations each correlating *energy* with one or more independent variables. Valid mathematical techniques must be used to derive the adjustment method for each *M&V Plan*. See Appendix B for some guidance on assessing the validity of mathematical methods.

and

- **Non-Routine Adjustments** – for those energy-governing factors which are not usually expected to change, such as: the facility size, the design and operation of installed equipment, the number of weekly production shifts, or the type of occupants. These static factors must be monitored for change throughout the reporting period. See Chapter 8.2 for discussion of non-routine adjustments.

Static Factors

Examples of *static factors* needing *non-routine adjustments* are changes in the:

- amount of space being heated or air conditioned,
- type of products being produced or number of production shifts per day
- building envelope characteristics (new insulation, windows, doors, air tightness),
- amount, type or use of the *facility's* and the users' equipment,
- indoor environmental standard (e.g. light levels, temperature, ventilation rate), and
- occupancy type or schedule.

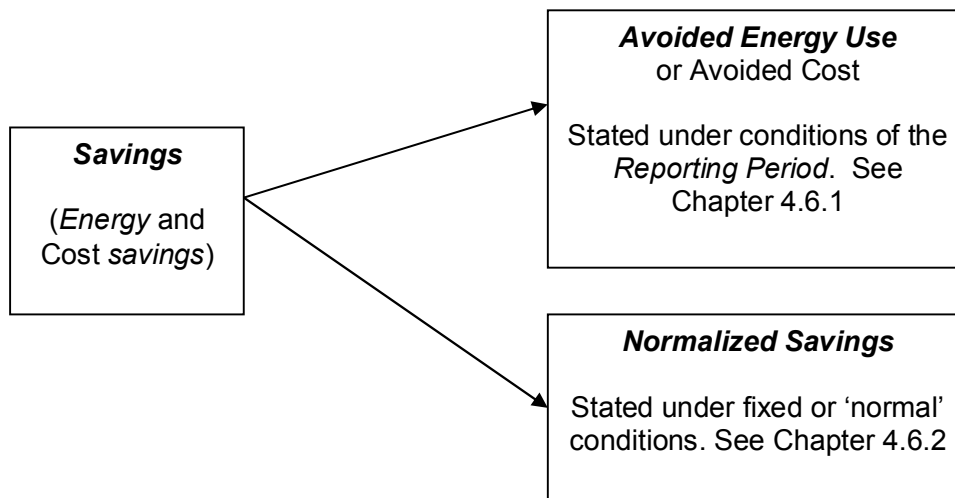
Therefore Equation 1) can be expressed more fully as:

$$\text{Savings} = (\text{Baseline Energy} - \text{Reporting-Period Energy}) \pm \text{Routine Adjustments} \pm \text{Non-Routine Adjustments} \quad 1a)$$

The *adjustments* terms in Equation 1a) are used to express both pieces of measured energy data under the same set of conditions. The mechanism of the adjustments depends upon whether *savings* are to be reported on the basis of the conditions of the *reporting period*, or normalized to some other fixed set of conditions as discussed below⁹.

⁹ The following general methods can be applied to Options A, B and C described in the rest of the Chapter 4. Option D generally includes the adjustments within the simulation, though the set of conditions for adjustment must still be chosen.

**Figure 3 Two
Types of
Savings**



Reporting-Period Basis or Avoided Energy Use

When *savings* are reported under the conditions of the *reporting period*, they can also be called *avoided energy use* of the *reporting period*. *Avoided energy use* quantifies *savings* in the *reporting period* relative to what *energy use* would have been without the *ECM(s)*.

When reporting *savings* under *reporting-period* conditions, *baseline-period energy* needs to be adjusted to *reporting-period* conditions.

For this common style of *savings* reporting Equation 1a) can be restated as:

$$\begin{aligned} \text{Avoided Energy Use (or Savings)} = & \\ & (\text{Baseline Energy} \pm \text{Routine Adjustments to reporting-period conditions} \\ & \pm \text{Non-Routine Adjustments to reporting-period conditions}) - \text{Reporting-Period Energy} \end{aligned}$$

This equation is often simplified to the following:

$$\begin{aligned} \text{Avoided Energy Use (or Savings)} = & \\ & \text{Adjusted-Baseline Energy} - \text{Reporting-Period Energy} \\ & \pm \text{Non-Routine Adjustments of baseline energy to reporting-period conditions} \end{aligned} \quad 1b)$$

Where *Adjusted-Baseline Energy* is defined as the *baseline energy* plus any *routine adjustments* needed to adjust it to the conditions of the *reporting period*.

The *adjusted-baseline energy* is normally found by first developing a mathematical model which correlates actual *baseline energy* data with appropriate *independent variable(s)* in the *baseline period*. Each *reporting period's independent variable(s)* are then inserted into this *baseline* mathematical model to produce the *adjusted-baseline energy use*.

Independent Variables

An *independent variable* is a parameter that is expected to change regularly and have a measurable impact on the *energy* use of a system or *facility*. For example, a common *independent variable* governing building energy use is outdoor temperature. Likewise in a manufacturing plant the number of units produced in a period is often an *independent variable* significantly affecting energy use. Another common *independent variable* is the number of seconds, hours or days in each metering period. See also Chapter 4.9.3.

Fixed Conditions Basis or *Normalized Savings*

Conditions other than those of the *reporting period* may be used as the basis for adjustment. The conditions may be those of the *baseline period*, some other arbitrary period, or a typical, average or 'normal' set of conditions.

Adjustment to a fixed set of conditions reports a style of *savings* which could be called "*normalized savings*" of the *reporting period*. In this method *energy* of the *reporting period* and possibly of the *baseline* period are adjusted from their actual conditions to the common fixed (or 'normal') set of conditions selected.

Equation 1c) restates the more general Equation 1a) for such *normalized savings* reports:

Normalized Savings =

$$\begin{aligned}
 & (\text{Baseline Energy} \pm \text{Routine Adjustments to fixed conditions} \\
 & \pm \text{Non-Routine Adjustments to fixed conditions}) \\
 & - (\text{Reporting Period Energy} \pm \text{Routine Adjustments to fixed conditions} \\
 & \pm \text{Non-Routine Adjustments to fixed conditions})
 \end{aligned}
 \tag{1c)$$

The calculation of the *reporting period routine-adjustments* term usually involves the development of a mathematical model correlating *reporting-period energy* with the *independent variables* of the *reporting period*. This model is then used to adjust *reporting-period energy* to the chosen fixed conditions. Further, if the fixed set of conditions is not from the *baseline period*, a mathematical model of *baseline energy* is also used to adjust *baseline energy* to the chosen fixed conditions.

What Basis for Adjustment, or Which Type of ‘Savings?’

Factors to consider when choosing between *avoided energy use* and *normalized savings*:

“*Avoided Energy Use*” style of *savings* (Equation 1b):

- are dependent upon the *reporting period*’s operating conditions. Even though *savings* can be properly adjusted for phenomena such as weather, the level of reported *savings* depends upon the actual weather.
- cannot be directly compared with *savings* predicted under *baseline* conditions.

“*Normalized savings*” style of *savings* (Equation 1c):

- are unaffected by *reporting-period* conditions since the fixed set of conditions are established once and not changed.
- can be directly compared with *savings* predicted under the same set of fixed conditions.
- can only be reported after a full cycle of *reporting-period energy* use, so that the mathematical correlation between *reporting-period energy* and operating conditions can be derived.

4.6 Overview Of IPMVP Options

The *energy* quantities in the several forms of Equation 1) can be measured by one or more of the following techniques:

- Utility or fuel supplier invoices, or reading utility meters and making the same adjustments to the readings that the utility makes.
- Special meters isolating an *ECM* or portion of a *facility* from the rest of the *facility*. Measurements may be periodic for short intervals, or continuous throughout the *baseline* or *reporting periods*.
- Separate measurements of parameters used in computing *energy* use. For example, equipment operating parameters of electrical load and operating hours can be measured separately and multiplied together to compute the equipment’s energy use.
- Measurement of proven *proxies* for *energy* use. For example, if the energy use of a motor has been correlated to the output signal from the variable speed drive controlling the motor, the output signal could be a proven *proxy* for motor energy.
- Computer simulation that is calibrated to some actual performance data for the system or *facility* being modeled. One example of computer simulation is DOE-2 analysis for buildings (Option D only).

If the a *energy* value is already known with adequate accuracy or when it is more costly to measure than justified by the circumstances, then measurement of *energy* may not be necessary or appropriate. In these cases, *estimates* may be made of some *ECM* parameters, but others must be measured (Option A only).

IPMVP provides four Options for determining *savings* (A, B, C and D). The choice among the Options involves many considerations including the location of the *measurement boundary* (see Chapter 4.4). If it is decided to determine *savings* at the *facility* level, Option C or D may be favored. However if only the performance of the *ECM* itself is of concern, a retrofit-isolation technique may be more suitable (Option A, B or D).

Table 2 summarizes the four Options that are detailed in Chapters 4.8 through 4.10. Examples of the use of the Options are contained in Appendix A. Section 4.11 offers guidance on selecting the proper Option for any specific project.

IPMVP Option	How Savings Are Calculated	Typical Applications
A. Retrofit Isolation: Key Parameter Measurement <i>Savings</i> are determined by field measurement of the key performance parameter(s) which define the <i>energy</i> use of the <i>ECM</i>'s affected system(s) and/or the success of the project. Measurement frequency ranges from short-term to continuous, depending on the expected variations in the measured parameter, and the length of the <i>reporting period</i>. Parameters not selected for field measurement are <i>estimated</i>. <i>Estimates</i> can be based on historical data, manufacturer's specifications, or engineering judgment. Documentation of the source or justification of the <i>estimated</i> parameter is required. The plausible <i>savings</i> error arising from <i>estimation</i> rather than measurement is evaluated.	Engineering calculation of <i>baseline</i> and <i>reporting period energy</i> from: ○ short-term or continuous measurements of key operating parameter(s); and ○ <i>estimated</i> values. <i>Routine and non-routine adjustments</i> as required.	A lighting retrofit where power draw is the key performance parameter that is measured periodically. Estimate operating hours of the lights based on <i>facility</i> schedules and occupant behavior.
B. Retrofit Isolation: All Parameter Measurement <i>Savings</i> are determined by field measurement of the <i>energy</i> use of the <i>ECM</i>-affected system. Measurement frequency ranges from short-term to continuous, depending on the expected variations in the <i>savings</i> and the length of the <i>reporting period</i>.	Short-term or continuous measurements of <i>baseline</i> and <i>reporting-period energy</i>, and/or engineering computations using measurements of proxies of <i>energy</i> use. Routine and non-routine adjustments as required.	Application of a variable-speed drive and controls to a motor to adjust pump flow. Measure electric power with a kW meter installed on the electrical supply to the motor, which reads the power every minute. In the <i>baseline period</i> this meter is in place for a week to verify <i>constant</i> loading. The meter is in place throughout the <i>reporting period</i> to track variations in power use.

Table 2
Overview of
IPMVP
Options

Table 2
Overview of
IPMVP
Options

IPMVP Option	How Savings Are Calculated	Typical Applications
C. Whole Facility <i>Savings</i> are determined by measuring energy use at the whole <i>facility</i> or sub- <i>facility</i> level. Continuous measurements of the entire <i>facility's energy</i> use are taken throughout the <i>reporting period</i> .	Analysis of whole <i>facility baseline</i> and <i>reporting period</i> (utility) meter data. <i>Routine adjustments</i> as required, using techniques such as simple comparison or regression analysis. <i>Non-routine adjustments</i> as required.	Multifaceted energy management program affecting many systems in a <i>facility</i> . Measure energy use with the gas and electric utility meters for a twelve month <i>baseline period</i> and throughout the <i>reporting period</i> .
D. Calibrated Simulation <i>Savings</i> are determined through simulation of the <i>energy</i> use of the whole <i>facility</i> , or of a sub- <i>facility</i> . Simulation routines are demonstrated to adequately model actual <i>energy</i> performance measured in the <i>facility</i> . This Option usually requires considerable skill in calibrated simulation.	Energy use simulation, calibrated with hourly or monthly utility billing data. (Energy end use metering may be used to help refine input data.)	Multifaceted energy management program affecting many systems in a facility but where no meter existed in the <i>baseline</i> period. Energy use measurements, after installation of gas and electric meters, are used to calibrate a simulation. <i>Baseline</i> energy use, determined using the calibrated simulation, is compared to a simulation of <i>reporting period</i> energy use.

4.7 Options A & B: Retrofit Isolation

Chapter 4.4 defines the concept of a *measurement boundary* encompassing the retrofitted equipment. Retrofit isolation allows the narrowing of the *measurement boundary* in order to reduce the effort required to monitor *independent variables* and *static factors*, when retrofits affect only a portion of the *facility*. However boundaries smaller than the total *facility* usually require additional meters at the *measurement boundary*. Narrow *measurement boundaries* also introduce the possibility of 'leakage' through unmeasured *interactive effects*.

Since measurement is of less than the total *facility*, the results of retrofit isolation techniques cannot be correlated to the *facility's* total *energy* use shown on utility bills. *Facility* changes beyond the *measurement boundary* but unrelated to the *ECM* will not be reported by retrofit-isolation techniques but will be included in the utility's metered consumption or demand.

Two Options are presented for isolating the *energy* use of the equipment affected by an *ECM* from the energy use of the rest of the facility:

- Option A: Retrofit Isolation: Key Parameter Measurement (See Chapter 4.8.1)
- Option B: Retrofit Isolation: All Parameter Measurement (See Chapter 4.8.2)

Isolation metering is placed at the *measurement boundary* between equipment which the *ECM* affects and equipment which it does not affect.

Retrofit Isolation Example

A boiler is replaced with a more efficient one. A *measurement boundary* is drawn around just the boiler so that the assessment of the new boiler is unaffected by variations in the heating load of the whole facility.

Meters for fuel consumption and boiler heat output are all that are needed to assess the efficiencies of the two boilers over their full range of operations. *Savings* are reported for the boiler retrofit by applying the observed efficiency improvement to an estimated annual boiler load. The boiler efficiency test is repeated annually.

When drawing a *measurement boundary*, care should be taken to consider any energy flows affected by the *ECM* but beyond the boundary. A method must be derived for estimating such *interactive effects* (See Chapter 4.4). For example, a lighting load reduction often reduces HVAC system energy use, but the only reasonable *measurement boundary* would encompass just the electricity use of the lights, not their heating and cooling energy impacts. In this case the *ECM's* effect on HVAC energy requirements is an *interactive effect*, which must be assessed. If the *interactive effect* is expected to be significant, engineering estimates could be made of the *interactive effect* as some fraction of the measured lighting-energy *savings*. Conventional heating and cooling calculations would be used to determine the appropriate fraction(s) for each season. However if the *measurement boundary* can be expanded to encompass *interactive effects*, there is no need to estimate them.

Apart from small estimated *interactive effects*, the *measurement boundary* defines the metering points and the scope of any *adjustments*, which may be used in the various forms of Equation 1). Only changes to *energy* systems and operating variables within the *measurement boundary* must be monitored to prepare the *adjustments* term(s) of Equation 1).

Chapter 4.5 discusses measurement periods, generally. Parameters may be continuously measured or periodically measured for short periods. The expected amount of variation in the parameter will govern the decision of whether to measure continuously or periodically. Where a parameter is not expected to change it may be measured immediately after *ECM* installation and checked occasionally throughout the *reporting period*. The frequency of this checking can be determined by beginning with frequent measurements to verify that the parameter is *constant*. Once proven *constant*, the frequency of measurement may be reduced. To maintain control on *savings* as measurement frequency drops, more frequent inspections or other tests might be undertaken to verify proper operations.

Continuous metering provides greater certainty in reported *savings* and more data about equipment operation. This information can be used to improve or optimize the operation of the equipment on a real-time basis, thereby improving the benefit of the *ECM* itself. Results from several studies have shown five to fifteen percent annual energy *savings* can be achieved through careful use of continuous data logging (Claridge et al. 1994, 1996; Haberl et al. 1995).

If measurement is not continuous and meters are removed between readings, the location of the measurement and the specifications of the measurement device should be recorded in the M&V Plan, along with the procedure for calibrating the meter being used. Where a parameter is expected to be constant, measurement intervals can be short and occasional. Electric motors in

an industrial plant provide a common example of *constant* power flow, assuming they have a *constant* load. However motor-operating periods may vary with the type of product being produced from day to day. Where a parameter may change periodically, the occasional measurements of the parameter (operating hours in this motor example) should happen at times representative of the normal system behavior.

Where a parameter may vary daily or hourly, as in most building heating or cooling systems, continuous metering may be simplest. For weather dependent loads, measurements may be taken over a long enough period to adequately characterize the load pattern through all parts of its normal annual cycle (i.e. each season, and weekday/weekend) and repeated as necessary through the *reporting period*. Examples of such day-type profiling can be found in Katipamula and Haberl (1991), Akbari et al. (1988), Hadley and Tomich (1986), Bou Saada and Haberl (1995a, 1995b) and Bou Saada et al. (1996).

Where multiple versions of the same *ECM* installation are included within the *measurement boundary*, statistically valid samples may be used as valid measurements of the total parameter. Such situations may arise, for example, where total lighting power draw cannot be read at the electrical panel due to the presence of non-lighting loads on the same panel. Instead a statistically significant sample of fixtures is measured before and after retrofit to assess the change in power draw. These sample data may be used as the 'measurements' of total lighting power draw. Appendix B-3 discusses the statistical issues involved in sampling.

Portable meters may be used if only short-term metering is needed. The costs of portable meters can be shared with other objectives. However, permanently installed meters also provide feedback to operating staff or automated control equipment for optimization of systems. Added meters may also enable billing of individual users or departments in the facility.

Retrofit isolation techniques are best applied where:

- Only the performance of the systems affected by the *ECM* is of concern, either due to the responsibilities assigned to the parties in an *energy performance contract*, or due to the *savings* of the *ECM* being too small to be detected in the time available using Option C.
- *Interactive effects* of the *ECM* on the *energy* use of other *facility* equipment can be reasonably estimated, or assumed to be insignificant.
- Possible changes to the *facility*, beyond the *measurement boundary*, would be difficult to identify or assess.
- The *independent variables*, which affect *energy* use, are not excessively difficult or expensive to monitor.
- Sub-meters already exist to isolate *energy* use of systems.
- Meters added at the *measurement boundary* can be used for other purposes such as operational feedback or tenant billing.
- Measurement of parameters is less costly than Option D simulations or Option C *non-routine adjustments*.
- Long term testing is not warranted.
- There is no need to directly reconcile savings reports with changes in payments to energy suppliers.

The unique characteristics of each of the retrofit-isolation techniques are discussed in Chapters 4.7.1 and 4.7.2, below. Common measurement issues arising when using retrofit-isolation techniques are discussed in Chapter 4.8.3..

4.7.1 Option A: Retrofit Isolation: Key Parameter Measurement

Under Option A, Retrofit Isolation: Key Parameter Measurement, *energy* quantities in Equation 1) can be derived from a computation using a combination of measurements of some parameters and *estimates* of the others. Such *estimates* should only be used where it can be shown that the combined uncertainty from all such *estimates* will not significantly affect the overall reported *savings*. Decide which parameters to measure and which to *estimate* by considering each parameter's contribution to the overall uncertainty of the reported *savings*. The *estimated values* and analysis of their significance should be included in the *M&V Plan* (Chapter 5). *Estimates* may be based on historical data such as recorded operating hours from the *baseline*, equipment manufacturer's published ratings, laboratory tests, or typical weather data.

If a parameter, such as hours of use is known to be *constant* and not expected to be impacted by the *ECM*, then its measurement in the *reporting period* is sufficient. The *reporting period* measurement of such constant parameter can also be considered a measurement of its *baseline* value.

Wherever a parameter, known to vary independently, is not measured in the facility during both the *baseline* and *reporting periods*, the parameter should be treated as an *estimate*.

Engineering calculations or mathematical modeling may be used to assess the significance of the errors in estimating any parameter in the reported *savings*. For example if a piece of equipment's operating hours are to be estimated, but may range from 2,100 to 2,300 hours per year, the calculated *savings* at 2,100 and 2,300 hours should be computed and the difference evaluated for its significance to the expected *savings*. The combined effect of all such possible *estimations* should be assessed before determining whether sufficient measurement is in place. See also Appendix B-5.1.

The selection of which factor(s) to measure may also be considered relative to the objectives of the project or the duties of a contractor undertaking some *ECM*-performance risk. Where a factor is significant to assessing performance, it should be measured. Other factors beyond the contractor's control can be estimated.

If a *savings* computation involves subtracting a measured parameter from an *estimated*

parameter, the result is an *estimate*. For example if a parameter is measured in the *reporting period* and subtracted from an unmeasured value for the same parameter in the *baseline period*, the resultant difference is only an *estimate*.

An example application of Option A is an *ECM* involving the installation of high-efficiency light fixtures, without changing lighting periods. *Savings* can be determined using Option A by metering the lighting-circuit power draw before and after retrofit while *estimating* the operating period. Other variations on this type of *ECM*, shown in

What to Measure?

Consider the example of a lighting project where *reporting period* power draw is measured, but *baseline* power is not measured. Therefore power draw should be treated as an *estimate* in designing an Option A procedure.

As a result, operating hours must be measured if the procedure is to adhere to IPMVP Option A.

Table 2 below, show the circumstances where *estimates* adhere to the guidance of Option A.

Table 2
Example
Lighting

Situation	Measurement vs. Estimation Strategy		Adherent to Option A?
	Operating Hours	Power Draw	
ECM reduces operating hours	Measure	<i>Estimate</i>	Yes
	<i>Estimate</i>	Measure	No
ECM reduces power draw	<i>Estimate</i>	Measure	Yes
	Measure	<i>Estimate</i>	No
ECM reduces both power draw and operating hours:			
Baseline power uncertain, operating hours known	<i>Estimate</i>	Measure	Yes
	Measure	<i>Estimate</i>	No
Power known but operating hours uncertain	Measure	<i>Estimate</i>	Yes
	<i>Estimate</i>	Measure	No
Power and operating hours poorly known	Measure	<i>Estimate</i>	No – Use Option B
	<i>Estimate</i>	Measure	

When planning an Option A procedure, consider both the amount of variation in *baseline energy* and the *ECM's energy* impact before establishing which parameter(s) to measure. The following three examples show the range of scenarios that may arise.

- ECM reduces a *constant* load without changing its operating hours. Example: industrial-plant lighting fixtures are replaced with more efficient ones, but the operating hours of the lights do not change. To reasonably measure the effect of the project, fixture power levels should be measured in the *baseline* and *reporting periods*, while operating hours are *estimated* in the energy calculations.
- ECM reduces operating hours while load is unchanged. Example: automatic controls shut down air compressors during unoccupied periods. To reasonably measure the effect of the project, compressors' operating time should be measured in both the *baseline* and *reporting periods*, while compressors' power can be estimated in the energy calculations.
- ECM reduces both equipment load and operating hours. Example: Resetting of temperature on a hot-water radiation system reduces overheating and induces occupants to close windows, thereby reducing boiler load and operating periods. When both load and operating periods are variable and uncertain, Option A cannot be used.

Generally, conditions of variable load or variable operating hours require more rigorous measurement and computations.

Option A: Calculations

General Equation 1) in Chapter 4.1 is used in all IPMVP adherent computations. However under Option A, there may be no need for *adjustments*, *routine* or *non-routine*, depending upon the location of the *measurement boundary*, the nature of any *estimated values*, the length of the *reporting period*, or the amount of time between *baseline* measurements and *reporting-period* measurements.

Similarly *baseline* or *reporting-period energy* measurements involve measurement of only one parameter under Option A, and estimation of the other. Therefore Equation 1) may simplify down to:

$$\text{Option A Savings} = \text{Estimated Value} \times (\text{Baseline-Period, measured parameter} - \text{Reporting-period, measured parameter}) \quad 1d)$$

Option A: Installation Verification

Since some values may be *estimated* under Option A, great care is needed to review the engineering design and installation to ensure that the *estimates* are realistic, achievable, and based on equipment that should truly produce *savings* as intended.

At defined intervals during the *reporting period*, the installation should be re-inspected to verify continued existence of the equipment and its proper operation and maintenance. Such re-inspections will ensure continuation of the potential to generate predicted *savings* and validate *estimated parameters*. The frequency of these re-inspections is determined by the likelihood of performance changes. Such likelihood can be established through initial frequent inspections to establish the stability of equipment existence and performance.

An example of a situation needing routine re-inspection is a lighting retrofit. You can determine *savings* by sampling of the performance of fixtures and counting the number of operating fixtures. In this case, the continued existence of the fixtures and operation of the lamps is critical to the *savings* determination. Similarly, where the settings of controls are assumed, but subject to tampering, regular recordings of control settings or actual equipment functions can limit the uncertainty of the *estimated values*.

Option A: Cost

Savings determinations under Option A can be less costly than under other Options, since the cost of *estimating* a parameter is often significantly less than the cost of measurement. However in some situations where *estimation* is the only possible route, a good *estimate* may be costlier than if direct measurement were possible. Cost planning for Option A should consider all elements: analysis, *estimation*, meter installation, and the ongoing cost to read and record data.

Option A: Best Applications

In addition to the retrofit isolation best applications in Chapter 4.8, above, Option A is best applied where:

- *Estimation* of non-key parameters may avoid possibly difficult *non-routine adjustments* when future changes happen within the *measurement boundary*.
- The uncertainty created by *estimations* is acceptable.
- The continued effectiveness of the *ECM* can be assessed by simple routine re-testing of key parameters.
- *Estimation* of some parameters is less costly than measurement of them in Option B or simulation in Option D.
- The key parameter(s) used in computing *savings* can be readily identified. Key parameters are parameters used to judge a project's or contractor's performance.

4.7.2 Option B: Retrofit Isolation: All Parameter Measurement

Option B, Retrofit Isolation: All Parameter Measurement, requires measurement of all Equation 1) *energy* quantities, or all parameters needed to compute *energy*.

The *savings* created by most types of *ECMs* can be determined with Option B. However, the degree of difficulty and costs increase as metering complexity increases. Option B methods will generally be more difficult and costly than those of Option A. However, Option B will produce more certain results where load and/or *savings* patterns are variable. These additional costs may be justifiable if a contractor is responsible for all factors affecting *energy savings*.

Option B: Calculations

General Equation 1) in Chapter 4.1 is used in all IPMVP adherent computations. However under Option B, there may be no need for *adjustments*, *routine* or *non-routine*, depending upon the location of the *measurement boundary*, the length of the *reporting period*, or the amount of time between *baseline* and *reporting period* measurements. Therefore, for Option B, Equation 1 may simplify down to:

$$\text{Option B Savings} = \text{Baseline Energy} - \text{Reporting-Period Energy} \quad 1e)$$

Option B: Best Applications

In addition to the retrofit-isolation methods in Chapter 4.8, above, Option B is best applied where:

- Meters added for isolation purposes will be used for other purposes such as operational feedback or tenant billing.
- Measurement of all parameters is less costly than simulation in Option D.
- *Savings* or operations within the *measurement boundary* are variable.

4.7.3 Retrofit Isolation Measurement Issues

Retrofit isolation usually requires the addition of special meters, on either a short term or permanent basis. These meters may be installed during an *energy* audit to help characterize *energy* use before design of the *ECM*. Or meters may be installed to measure *baseline* performance for an *M&V Plan*.

You can measure temperature, humidity, flow, pressure, equipment runtime, electricity or thermal energy, for example, at the *measurement boundary*. Follow good measurement practices to enable calculation of *energy savings* with reasonable accuracy and repeatability. Measurement practices are continually evolving as metering equipment improves. Therefore, use the latest measurement practices to support your *savings* (see also Chapter 8.11).

The following sections define some key measurement issues to consider when using retrofit-isolation techniques.

Electricity Measurements

To measure electricity accurately we measure the voltage, amperage and power factor, or true rms¹⁰ wattage with a single instrument. However measurement of amperage and voltage alone can adequately define wattage in purely resistive loads, such as incandescent lamps and resistance heaters without blower motors. When measuring power, make sure that a resistive load's electrical wave-form is not distorted by other devices in the *facility*.

Measure electric demand at the same time that the power company determines the peak demand for its billing. This measurement usually requires continuous recording of the demand at the sub-meter. From this record, the sub-meter's demand can be read for the time when the power company reports that the peak demand occurred on its meter. The power company may reveal the time of peak demand either on its invoices or by special report.

Electric-demand measurement methods vary amongst utilities. The method of measuring electric demand on a sub-meter should replicate the method the power company uses for the relevant billing meter. For example, if the power company calculates peak demand using fixed

¹⁰ Rms (root mean squared) values can be reported by solid state digital instruments to properly account for the net power when wave distortions exist in alternating current circuits.

15 minute intervals, then the recording meter should be set to record data for the same 15 minute intervals. However if the power company uses a moving interval to record electric demand data, the data recorder should have similar capabilities. Such moving interval capability can be emulated by recording data on one minute fixed intervals and then recreating the power company's intervals using post-processing software. However, care should be taken to ensure that the facility does not contain unusual combinations of equipment that generate high one minute peak loads which may show up differently in a moving interval than in a fixed interval. After processing the data into power company intervals, convert it to hourly data for archiving and further analysis.

Calibration

Meters should be calibrated as recommended by the equipment manufacturer, and following procedures of recognized measurement authorities. Primary standards and no less than third-order-standard traceable calibration equipment should be utilized wherever possible. Sensors and metering equipment should be selected based in part on the ease of calibration and the ability to hold calibration. An attractive solution is the selection of equipment that is self-calibrating.

Selected calibration references are provided in Chapter 10.3.

4.8 Option C: Whole Facility

Option C: Whole Facility, involves use of utility meters, whole-*facility* meters, or sub-meters to assess the energy performance of a total *facility*. The *measurement boundary* encompasses either the whole *facility* or a major section. This Option determines the collective *savings* of all *ECMs* applied to the part of the *facility* monitored by the *energy* meter. Also, since whole-*facility* meters are used, *savings* reported under Option C include the positive or negative effects of any non-*ECM* changes made in the *facility*.

Option C is intended for projects where expected *savings* are large compared to the random or unexplained energy variations which occur at the whole-*facility* level. If savings are large compared to the unexplained variations in the *baseline-energy* data, then identifying *savings* will be easy. Also the longer the period of *savings* analysis after *ECM* installation, the less significant is the impact of short-term unexplained variations¹¹. Typically *savings* should exceed 10% of the *baseline energy* if you expect to confidently discriminate the *savings* from the *baseline* data when the *reporting period* is shorter than two years.

Identifying *facility* changes that will require *non-routine adjustments* is the primary challenge associated with Option C, particularly when *savings* are monitored for long periods. (See also Chapter 8.2 on *non-routine baseline adjustments*.) Therefore, you should perform periodic inspections of all equipment and operations in the *facility* during the *reporting period*. These inspections identify changes in the *static factors* from *baseline* conditions. Such inspections may be part of regular monitoring to ensure that the intended operating methods are still being followed.

4.8.1 Option C: Energy Data Issues

Where utility supply is only measured at a central point in a group of *facilities*, sub-meters are needed at each *facility* or group of *facilities* for which individual performance is assessed.

Several meters may be used to measure the flow of one *energy* type into a facility. If a meter supplies *energy* to a system that interacts with other *energy* systems, directly or indirectly, this meter's data should be included in the whole-*facility savings* determination.

¹¹ See Appendix B-5. ASHRAE (2002) provides quantitative methods for assessing the impact of variations in the *baseline* data as the *reporting period* lengthens.

Meters serving non-interacting *energy* flows, for which *savings* are not to be determined, can be ignored. Separately metered outdoor-lighting circuits is one example.

Determine *savings* separately for each meter or sub-meter serving a *facility* so that performance changes can be assessed for separately metered parts of the *facility*. However, where a meter measures only a small fraction of one *energy* type's total use, it may be totaled with the larger meter(s) to reduce data-management tasks. When electrical meters are combined this way, it should be recognized that small consumption meters often do not have demand data associated with them so that the totalized consumption data will no longer provide meaningful load factor information.

If several different meters are read on separate days, then each meter having a unique billing period should be separately analyzed. The resultant *savings* can be combined after analysis of each individual meter, if the dates are reported.

If any of the energy data are missing from the *reporting period*, a *reporting-period* mathematical model can be created to fill in missing data. However the reported *savings* for the missing period should identify these *savings* as "missing data."

4.8.2 Option C: Energy Invoices Issues

Energy data for Option C are often derived from utility meters, either through direct reading of the meter, or from utility invoices. Where utility bills are the source of data, it should be recognized that a utility's need for regular meter reading is not usually as great as the needs of M&V. Utility bills sometimes contain estimated data, especially for small accounts. Sometimes it cannot be determined from the bill itself whether the data came from an estimate or an actual meter reading. Unreported estimated meter readings create unknown errors for estimated month(s) and also for the subsequent month of the actual meter reading. However the first invoice with an actual reading after one or more estimates will correct the previous errors in energy quantities. *Savings* reports should note when estimates are part of the utility data.

When an electrical utility estimates a meter reading, no valid data exist for the electrical demand of that period.

Energy may be supplied indirectly to a facility, through on-site storage facilities, such as for oil, propane or coal. In these situations, the energy supplier's shipment invoices do not represent the facility's actual consumption during the period between shipments. Ideally a meter downstream of the storage facility measures energy use. However where there is no downstream meter, inventory-level adjustments for each invoice period should supplement the invoices.

4.8.3 Option C: Independent Variables

Regularly changing parameters affecting a *facility's energy* use, are called *independent variables* (see also box in Chapter 4.6.1). Common *independent variables* are weather, production volume and occupancy. Weather has many dimensions, but for whole-facility analysis, weather is often just outdoor dry-bulb temperature. Production has many dimensions, depending upon the nature of the industrial process. Production is typically expressed in mass units or volumetric units of each product. Occupancy is defined in many ways, such as hotel-room occupancy, office-building occupancy hours, occupied days (weekdays/weekends), or restaurant-meal sales.

Mathematical modeling can assess *independent variables* if they are cyclical. *Regression analysis* and other forms of mathematical modeling can determine the number of *independent variables* to consider in the *baseline* data (See Appendix B-2). Parameters, which have a

significant effect on the *baseline energy* use, should be included in the *routine adjustments* when determining *savings*¹² using Equation 1a), b) or c).

Independent variables should be measured and recorded at the same time as the *energy* data. For example, weather data should be recorded daily so they can be totaled to correspond with the exact monthly energy-metering period, which may be different from the calendar month. Use of simple monthly mean temperature data for a non-calendar month energy metering period introduces unnecessary error into the analysis.

4.8.4 Option C: Calculations and Mathematical Models

For Option C, the *routine adjustments* term of Equation 1a) is calculated by developing a valid mathematical model of each meter's *energy*-use pattern. A model may be as simple as an ordered list of twelve measured monthly *energy* quantities without any adjustments. However a model often includes factors derived from *regression analysis*, which correlate *energy* to one or more *independent variables* such as outdoor temperature, *degree days*, metering period length, production, occupancy, and/or operating mode. Models can also include a different set of regression parameters for each range of conditions, such as summer or winter in buildings with seasonal *energy*-use variations. For example, in schools where the building's *energy* use differs between the school year and the vacation period, you may need separate regression models for the different usage periods (Landman and Haberl 1996a; 1996b).

Option C should use complete years (12, 24, or 36 months) of continuous data, during the *baseline period*, and continuous data during the *reporting periods* (Fels 1986). Models, which use other numbers of months, (9, 10, 13, or 18 months, for example) can create statistical bias by under or over-representing normal modes of operation.

Meter data can be hourly, daily or monthly whole-*facility* data. Hourly data should be combined into daily data to limit the number of *independent variables* required to produce a reasonable *baseline* model, without significantly increasing the uncertainty in computed *savings* (Katipamula 1996, Kissock et al. 1992). Variation in the daily data often results from the weekly cycle of most facilities.

Many mathematical models are appropriate for Option C. To select the one most suited to the application, consider statistical-evaluation indices, such as R^2 and t (see Appendix B-2.2)¹³. Appendix B-2.2 or published statistical literature can help you demonstrate the statistical validity of your selected model.

4.8.5 Option C: Metering

Whole-*facility energy* measurements can use the utility's meters. Utility-meter data is considered 100% accurate for determining *savings* because this data defines the payment for *energy*. Utility-meter data is usually required to meet commercial accuracy regulations for sale of *energy* commodities.

The *energy* supplier's meter(s) may be equipped or modified to provide an electrical pulse output that can be recorded by the *facility's* monitoring equipment. The *energy*-per-pulse constant of the pulse transmitter should be calibrated against a known reference such as similar data recorded by the utility meter.

¹² All other parameters affecting *energy* use (i.e. "*static factors*" see box in Chapter 4.6) should be measured and recorded in the *baseline* and *reporting periods* so that *non-routine adjustments* can be made if needed (see Chapter 8.8)

¹³ Additional information concerning these selection procedures can be found in Reynolds and Fels (1988), Kissock et al. (1992, 1994) and in the ASHRAE Handbook of Fundamentals (2005) Chapter 32. ASHRAE (2002) also provides several statistical tests to validate the usefulness of derived regression models.

Separate meters installed by the *facility* owner can measure whole-*facility energy*. The accuracy of these meters should be considered in the *M&V Plan*, together with a way of comparing its readings with the utility meter readings.

4.8.6 Option C: Cost

Option C's cost depends on the source of the *energy* data, and the difficulty of tracking *static factors* within the *measurement boundary* to enable *non-routine adjustments* during the *reporting period*. The utility meter or an existing sub-meter works well if the meter's data is properly recorded. This choice requires no extra metering cost.

The cost of tracking changes in *static factors* depends on the *facility's* size, the likelihood of *static-factor* change, the difficulty of detecting changes, and the surveillance procedures already in place.

4.8.7 Option C: Best Applications

Option C is best applied where:

- The *energy* performance of the whole *facility* will be assessed, not just the *ECMs*.
- There are many types of *ECMs* in one *facility*.
- The *ECMs* involve activities whose individual *energy* use is difficult to separately measure (operator training, wall or window upgrades, for example).
- The *savings* are large compared to the variance in the *baseline* data, during the *reporting period* (See Appendix B-1.2).
- When Retrofit-Isolation techniques (Option A or B) are excessively complex. For example, when *interactive effects* or interactions between *ECMs* are substantial.
- Major future changes to the facility are not expected during the *reporting period*.
- A system of tracking *static factors* can be established to enable possible future *non-routine adjustments*.
- Reasonable correlations can be found between *energy* use and other *independent variables*.

4.9 Option D: Calibrated Simulation

Option D, Calibrated Simulation, involves the use of computer simulation software to predict *facility energy* for one or both of the terms in Equation 1). A simulation model must be "calibrated" so that it predicts an *energy* pattern that approximately matches actual metered data.

Option D may be used to assess the performance of all *ECMs* in a *facility*, akin to Option C. However, the Option D simulation tool allows you to also estimate the *savings* attributable to each *ECM* within a multiple-*ECM* project.

Option D may also be used to assess just the performance of individual systems within a *facility*, akin to Options A and B. In this case, the system's *energy* use must be isolated from that of the rest of the facility by appropriate meters, as discussed in Chapters 4.5.1 and 4.7.

Option D is useful where:

- *Baseline energy* data do not exist or are unavailable. Such situation may arise for:
 - A new construction project
 - A facility expansion needing to be assessed separately from the rest of the facility, or

- A centrally metered campus of *facilities* where no individual *facility* meter exists in the *baseline period*, but where individual meters will be available after *ECM* installation.
- *Reporting-period energy* data are unavailable or obscured by factors that are difficult to quantify. Sometimes it is too difficult to predict how future *facility* changes might affect *energy* use. Industrial-process changes or new equipment often make the computation of *non-routine adjustments* so inaccurate that Options A, B or C would create excessive error in the *savings* determination.
- It is desired to determine the *savings* associated with individual *ECMs*, but measurements with Options A or B are too difficult or costly.

If the *reporting-period energy* is predicted by the simulation software, the determined *savings* persist only if the simulated operating methods continue. Periodic inspections will identify changes from *baseline* conditions and modeled equipment performance (see also Chapter 4.7.1. Option A Installation Verification). Simulation runs should be adjusted accordingly.

Option D is the primary *M&V* approach for assessing *energy* efficiency inclusions in new *facility* designs. The IPMVP Volume III Part I, titled “Concepts and Options for Determining Savings In New Construction”, provides guidance on a variety of *M&V* techniques applicable to new buildings.

Accurate computer modeling and calibration to measured *energy* data are the major challenges associated with Option D. To control the costs of this method while maintaining reasonable accuracy, the following points should be considered when using Option D:

- Simulation analysis should be conducted by trained personnel who are experienced with both the software and the calibration techniques.
- Input data should represent the best available information including as much as possible of available actual performance data from key components in the facility
- The simulation inputs need to be adjusted so its results match both the demand and consumption data from monthly utility bills, within acceptable tolerances (i.e. “calibrated”). Close agreement between predicted and actual annual total energy is usually insufficient demonstration that the simulation adequately predicts the *energy* behavior of the facility (See Chapter 4.9.2).
- Option D requires careful documentation. Simulation printouts, survey data and the metering or monitoring data used to define input values and calibrate the simulation model should be kept in paper and electronic files. The version number should be declared of publicly available software, so that another person can review the computations.
- For new construction projects, modeling efforts can be streamlined by retaining the building energy modeler that created the “as-designed” model to create the calibrated, “as-built” and adjusted baseline models.

Building types which are not easily simulated include those with:

- large atriums,
- a significant fraction of the space underground or ground coupled,
- unusual exterior shapes,
- complex shading configurations, or
- a large number of distinct zones of temperature control.

Some building *ECMs* cannot be simulated without great difficulty, such as:

- addition of radiant barriers in attics, and

- some complex HVAC system changes.

4.9.1 Option D: Types of Simulation Programs

Information on building simulation programs in common use in different parts of the world can be found in Appendix C herein

Whole-building-simulation programs usually use hourly calculation techniques. However simplified HVAC system models may also be used if the building's heat losses, heat gains, internal loads, and HVAC systems are simple. Other types of special-purpose programs may be used to simulate *energy* use and operation of devices or industrial processes.

Any software used must be well documented and well understood by the user. Due to the wide variety of available methods, it is prudent to receive acceptance by the owner or project authority of the proposed modeling program before commencing analysis.

4.9.2 Option D: Calibration

Savings determined with Option D are based on one or more complex estimates of *energy* use. The accuracy of the *savings* depends on how well the simulation models actual equipment performance, and how well calibrated it is to metered *energy* performance.

Calibration is achieved by verifying that the simulation model reasonably predicts the *energy* patterns of the *facility* by comparing model results to a set of calibration data. These calibration data include measured *energy* data, *independent variables* and *static factor*.

Calibration of building simulations is usually done with 12 monthly utility bills. These bills should be from a period of stable operation. In a new building, it may take a number of months before full occupancy and before the staff learns the best ways to operate the facility. The calibration data should be documented in the *M&V Plan* along with a description of its sources.

Detailed operating data from the *facility* help to develop the calibration data. These data might include operating characteristics, occupancy, weather, loads and equipment efficiency. Some variables may be measured for short intervals (day, week or month) or extracted from existing operating logs. The accuracy of meters should be verified for critical measurements. If resources permit, building ventilation and infiltration should be measured because these quantities often vary widely from expectations. One-time measurements will improve simulation accuracy without much additional cost. On/off tests can measure lighting, receptacle loads and motor control centers. These tests can be performed over a weekend using a data logger or building automation system to record whole-*facility energy* use, usually at one-minute intervals. Sometimes, inexpensive portable loggers, which are synchronized to a common time, are also effective for short-term measurement (Benton et al. 1996, Houcek et al. 1993, Soebarto 1996).

Following collection of as much calibration data as possible, the steps in calibrating the simulation are as listed below.

1. Assume other necessary input parameters, and document them.
2. Whenever possible, gather actual weather data from the calibration period, especially if weather conditions varied significantly from standard-year weather data used in the basic simulations. However, obtaining and preparing actual weather data for use with a simulation may be time-consuming and expensive. If developing an actual weather data file is too difficult, then adjust an average weather file to resemble actual weather data using valid statistical methods.
3. Run the simulation and verify that it predicts operating parameters such as temperature and humidity.
4. Compare the simulated *energy* results with the metered *energy* data from the calibration period, on an hourly or monthly basis.

5. Evaluate patterns in the differences between simulation results and calibration data. Bar charts, monthly percent difference time-series graphs, and monthly x-y scatter plots help to identify the error patterns. The calibration accuracy should be established in the *M&V Plan* to accommodate the *M&V* budget.
6. Revise input data in step 1 and repeat steps 3 and 4 to bring predicted results within the calibration specifications in 5, above. Collect more actual operating data from the facility to meet the calibration specification if necessary.

The creation and calibration of a simulation is time consuming. Using monthly rather than hourly *energy* data helps to limit the effort needed for calibration. However if Option D will be used to determine savings at the ECM level, calibration of major end-uses, systems, and/or equipment impacted by the ECMs is recommended.

4.9.3 Option D: Calculations

Savings can be determined using calibrated simulation results representing the baseline energy and/or the *reporting-period* energy. For projects with a physical baseline, the two calibrated models include one with the *ECMs* and one without them. For projects with a hypothetical baseline, calibrated models may include the hypothetical baseline and the as-built (*reporting period*) conditions, but measured data will only be available for calibration under as-built conditions. In either case, both models and measured energy data must be under the same set of operating conditions.

Savings with Option D can be estimated using two forms of Equation 1), Equations 1f) and 1g)¹⁴. Both forms presume that the calibration ‘error’ equally affects both *baseline* and *reporting period* models. The same savings will be determined from the two equations for any given set of data and simulations.

$$\begin{aligned} \text{Savings} = & \text{Baseline energy from the calibrated model [hypothetical or without ECMs]} \\ & - \text{Reporting-period energy from the calibrated model [with ECMs]} \end{aligned} \quad 1f)$$

One of the model-derived energy terms in Equation 1f) may be replaced by the actual measured *energy*. However, you must adjust your calculations for the calibration error for each month in the calibration period, using Equation 1g):

$$\begin{aligned} \text{Savings} = & \text{Baseline energy from the calibrated model [hypothetical or without ECMs]} \\ & - \text{Actual calibration-period energy} \\ & +/- \text{Calibration error in the corresponding calibration reading} \end{aligned} \quad 1g)$$

Equation 1 may be most easily understood by non-technical persons, since the final savings computation uses actual metered data rather than only the results of simulation models.

4.9.4 Option D: Ongoing Savings Reporting

If multi-year performance evaluation is required, Option D may be used for the first year after the *ECMs* are installed. In later years, Option C may be less costly than Option D if you use as the *baseline* the meter data from the first year of steady operation after installation. Then Option C is used to determine whether *energy* use changes from the first year of operation after the *ECM* was installed. In this situation, the first year of steady operation’s *energy* use would be

¹⁴ Equations 1f) and 1g) are the same as Methods 1 and 2, respectively, found in IPMVP Vol III Part 1.

used: a) to calibrate an Option D simulation model, and b) to establish an Option C *baseline* for measuring additional *savings* (or losses) in the second year and beyond.

4.9.5 Option D: Best Applications

Option D is usually used where no other Option is feasible.

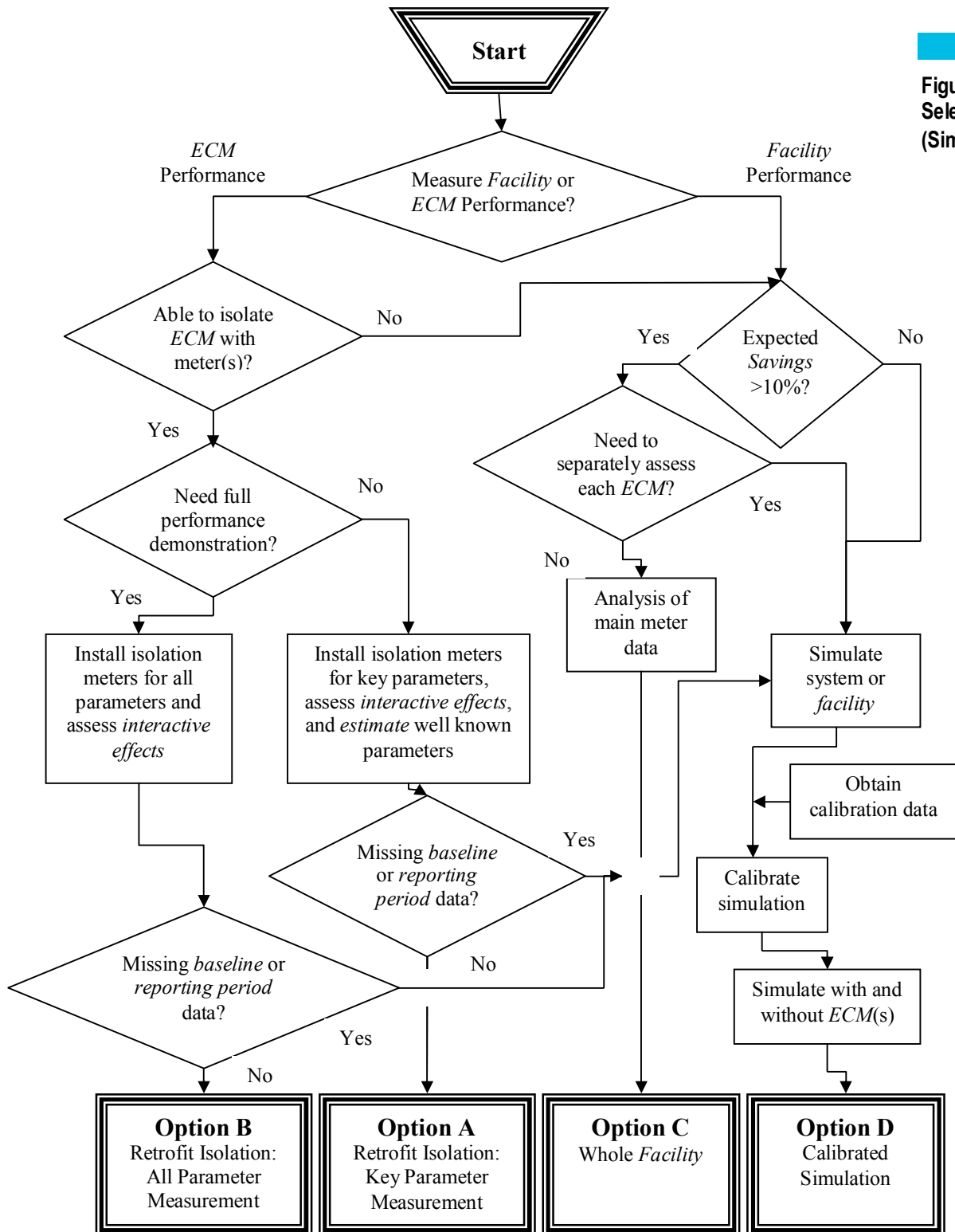
Option D is best applied where:

- Either *baseline energy* data or *reporting-period energy* data, but not both, are unavailable or unreliable.
- There are too many *ECMs* to assess using Options A or B.
- The *ECMs* involve diffuse activities, which cannot easily be isolated from the rest of the *facility*, such as operator training or wall and window upgrades.
- The performance of each *ECM* will be estimated individually within a multiple-*ECM* project, but the costs of Options A or B are excessive.
- Interactions between *ECMs* or *ECM interactive effects* are complex, making the isolation techniques of Options A and B impractical.
- Major future changes to the facility are expected during the *reporting period*, and there is no way to track the changes and/or account for their impact on *energy* use.
- An experienced *energy*-simulation professional is able to gather appropriate input data to calibrate the simulation model.
- The *facility* and the *ECMs* can be modeled by well documented simulation software.
- Simulation software predicts metered calibration data with acceptable accuracy.
- Only one year's performance is measured, immediately following installation and commissioning of the *energy*-management program.

4.10 Option Selection Guide

The selection of an IMPVP Option is a decision that is made by the designer of the *M&V* program for each project, based on the full set of project conditions, analysis, budgets and professional judgment. Figure 4 outlines common logic used in Option selection.

**Figure 4 Option
Selection Process
(Simplified)**



It is impossible to generalize on the best IPMVP Option for any type of situation. However some key project characteristics suggest commonly favored Options as shown in Table 3 below.

**Table 3 Suggested
(not the only) Options
- Marked by X**

ECM Project Characteristic	Suggested Option			
	A	B	C	D
Need to assess <i>ECMs</i> individually	X	X		X
Need to assess only total facility performance			X	X
Expected <i>savings</i> less than 10% of utility meter	X	X		X
Significance of some <i>energy</i> driving variables is unclear		X	X	X
<i>Interactive effects</i> of ECM are significant or unmeasurable			X	X
Many future changes expected within <i>measurement boundary</i>	X			X
Long term performance assessment needed	X		X	
Baseline data not available				X
Non-technical persons must understand reports	X	X	X	
Metering skill available	X	X		
Computer simulation skill available				X
Experience reading utility bills and performing regression analysis available			X	

4.11 Persistence of Savings

An organisation which has improved its energy efficiency is more at risk from adverse shifts in performance, not only because its ECMs may be of a kind that can fail, fade or be bypassed, but also because the additional headroom in its energy budgets may allow unrelated incidents of avoidable energy waste to pass unchallenged.

Persistence of energy savings can be achieved beyond the M&V reporting period by completing follow on efforts that build on M&V. One approach is “Monitoring and Targeting” (M&T), which can seamlessly follow the M&V process. If Options B and C (see sections 4.6 above) have been used for verifying savings, the project will have metering in place for the routine measurement of consumption. More importantly, models will also have been developed that correlate energy use with driving factors such as weather. These same models can be ‘re-tuned’ to estimate energy consumption that accounts for ECM installation. This enables a periodic comparison of actual and expected consumption, which will readily reveal and quantify any loss of ECM effect (or unrelated waste) enabling prompt remedial action to be taken in cases where the unexpected avoidable cost is deemed significant.

CHAPTER 5 M&V PLAN CONTENTS

The preparation of an *M&V Plan* is a recommended part of *savings* determination. Advance planning ensures that all data needed for *savings* determination will be available after implementation of the *ECM(s)*, within an acceptable budget.

Data from the *baseline* and details of the *ECMs* may be lost over time. Therefore record them for future reference in case conditions change or *ECMs* fail. Documentation should be easy to find and easy to understand by verifiers and others, because years may pass before these data are needed.

A complete *M&V Plan* should include discussion of the following 13 topics:

1. **ECM Intent** Describe the *ECM*, its intended result, and the *operational verification* procedures that will be used to verify successful implementation of each *ECM*. Identify any planned changes to conditions of the *baseline*, such as unoccupied building temperature settings.
2. **Selected IPMVP Option and Measurement Boundary** Specify which IPMVP Option, defined in Chapters 4.8 – 4.10, will be used to determine *savings*. This identification should include the date of publication or the version number and Volume number of the IPMVP edition being followed (IPMVP Volume I EVO 10000-1:2012), for example). Identify the *measurement boundary* of the *savings* determination. The boundary may be as narrow as the flow of *energy* through a pipe or wire, or as broad as the total *energy* use of one or many *facilities*. Describe the nature of any *interactive effects* beyond the *measurement boundary* together with their possible effects (see Chapter 4.4).
3. **Baseline: Period, Energy and Conditions** Document the facility's *baseline* conditions and *energy* data, within the *measurement boundary*. (In *energy performance contracts*, *baseline energy* and *baseline* conditions may be defined by either the owner or the *ESCO*, providing the other party is given adequate opportunity to verify them.)

An *energy audit* used for establishing the objectives of a *savings* program or terms of an *energy performance contract* usually provides most if not all of the *baseline* documentation needed in the *M&V Plan*. This *baseline* documentation should include:

- a) Identification of the *baseline period* (Chapter 4.5.1)
- b) All *baseline energy* consumption and demand data
- c) All *independent variable* data coinciding with the *energy* data (e.g. production rate, ambient temperature)
- d) All *static factors* coinciding with the *energy* data:
 - Occupancy type, density and periods
 - Operating conditions for each *baseline* operating period and season, other than the *independent variables*. (For example, in an industrial process, *baseline* operating conditions might include product type(s), raw material type, and number of production shifts per day. In a building *baseline* operating conditions might include light level, space temperature humidity and ventilation levels. An assessment of thermal comfort and/or indoor air quality (IAQ) may also prove useful in cases where the new system performs differently than the old inefficient system. See IPMVP Volume II.)
 - Description of any *baseline* conditions that fall short of required conditions. For example, the space is under-heated during the *baseline*, but the *ECM* will restore the desired temperature. Details of all adjustments that are necessary to the *baseline*

energy data to reflect the energy-management program's expected improvement from *baseline* conditions.

- Size, type, and insulation of any relevant building envelope elements such as walls, roofs, doors, windows.
- Equipment inventory: nameplate data, location, condition. Photographs or videotapes are effective ways to record equipment condition.
- Equipment operating practices (schedules and setpoints, actual temperatures and pressures)
- Significant equipment problems or outages during the *baseline period*.

The *baseline* documentation typically requires well-documented audits, surveys, inspections and/or short-term metering activities. The extent of this information is determined by the *measurement boundary* chosen or the scope of the *savings* determination.

Where whole-*facility* M&V methods are employed (Chapter 4.9 or 4.10), all *facility* equipment and conditions should be documented.

4. **Reporting Period** Identify the *reporting period*. This period may be as short as an instantaneous measurement during commissioning of an *ECM*, or as long as the time required to recover the investment cost of the *ECM* program (See Chapter 4.5.2).
5. **Basis for Adjustment** Declare the set of conditions to which all *energy* measurements will be adjusted. The conditions may be those of the *reporting period* or some other set of fixed conditions. As discussed in Chapter 4.6, this choice determines whether *savings* are reported as *avoided energy* (4.6.1) or as *normalized savings* (4.6.2).
6. **Analysis Procedure** Specify the exact data analysis procedures, algorithms and assumptions to be used in each *savings* report. For each mathematical model used, report all of its terms and the range of *independent variables* over which it is valid.
7. **Energy Prices** Specify the energy prices that will be used to value the *savings*, and whether and how savings will be adjusted if prices change in future (See Chapter 8.1).
8. **Meter Specifications** Specify the metering points, and period(s) if metering is not continuous. For non-utility meters, specify: meter characteristics, meter reading and witnessing protocol, meter commissioning procedure, routine calibration process, and method of dealing with lost data (see Chapter 8.11.1).
9. **Monitoring Responsibilities** Assign responsibilities for reporting and recording the *energy* data, *independent variables* and *static factors* within the *measurement boundary* during the *reporting period*.
10. **Expected Accuracy** Evaluate the expected accuracy associated with the measurement, data capture, sampling and data analysis. This assessment should include qualitative and any feasible quantitative measures of the level of uncertainty in the measurements and adjustments to be used in the planned *savings* report (See Chapter 8.3 and Appendix B).
11. **Budget** Define the budget and the resources required for the *savings* determination, both initial setup costs and ongoing costs throughout the *reporting period*.
12. **Report Format** Specify how results will be reported and documented (see Chapter 6). A sample of each report should be included.
13. **Quality Assurance** Specify quality-assurance procedures that will be used for *savings* reports and any interim steps in preparing the reports.

Depending upon the circumstances of each project, some additional specific topics should also be discussed in a complete *M&V Plan*:

For Option A:

- **Justification of Estimates** Report the values to be used for all *estimated values*. Explain the source of these *estimated values*. Show the overall significance of these *estimates* to the total expected *savings* by reporting the range of the possible *savings* associated with the range of plausible values of the *estimated* parameters.
- **Periodic Inspections** Define the periodic inspections that will be performed in the *reporting period* to verify that equipment is still in place and operating as assumed when determining the *estimated* values.

For Option D:

- **Software Name** Report the name and version number of the simulation software to be used.
- **Input/Output Data** Provide a paper and electronic copy of the input files, output files, and weather files used for the simulation.
- **Measured Data** Note which input parameters were measured and which were estimated. Describe the process of obtaining any measured data.
- **Calibration** Report the *energy* and operating data used for calibration. Report the accuracy with which the simulation results match the calibration *energy* data.

Where the nature of future changes can be anticipated, define methods for making the relevant *non-routine adjustments*.

Time and budget requirements (item 11, above) are often underestimated, which leads to incomplete data collection. A less accurate and less expensive *savings* determination is better than an incomplete or poorly done determination, which is theoretically more accurate but under-funded. Chapter 8.5 addresses cost/benefit tradeoffs.

Issues arising in developing *M&V Plans* are discussed in the examples shown in Appendix A. The website of Efficiency Valuation Organization (www.evo-world.org) contains a growing selection of sample *M&V Plans*.

CHAPTER 6 M&V REPORTING

M&V Reports should be prepared and presented as defined in the *M&V Plan* (Chapter 5)

Complete *M&V* reports should include at least:

- Observed data of the *reporting period*: the *measurement period* start and end points in time, the *energy* data, and the values of the *independent variables*
- Description and justification for any corrections made to observed data
- For Option A the agreed *estimated values*
- *Energy* price schedule used
- All details of any *baseline non-routine adjustment* performed. Details should include an explanation of the change in conditions since the *baseline period*, all observed facts and assumptions, and the engineering calculations leading to the adjustment.
- Computed *savings* in *energy* and monetary units.

M&V reports should be written to their readers' levels of understanding.

Energy managers should review the *M&V* reports with the *facility's* operating staff. Such reviews may uncover useful information about how the facility uses *energy*, or where operating staff could benefit from more knowledge of the *energy*-consumption characteristics of their *facility*.

CHAPTER 7 ADHERENCE WITH IPMVP

IPMVP is a framework of definitions and methods for properly assessing *savings* in energy or water use or demand. The IPMVP guides users in developing *M&V Plans* for specific projects. IPMVP is written to allow maximum flexibility in creating *M&V Plans*, while adhering to the principles of accuracy, completeness, conservativeness, consistency, relevance and transparency (Chapter 3).

Users claiming adherence with IPMVP must:

1. Identify the person responsible for approving the site-specific *M&V Plan*, and for making sure that the *M&V Plan* is followed for the duration of the *reporting period*.
2. Develop a complete *M&V Plan* which:
 - clearly states the date of publication or the version number of the IPMVP edition and Volume being followed,
 - uses terminology consistent with the definitions in the version of IPMVP cited,
 - includes all information mentioned in the *M&V Plan* chapter (Chapter 5 of the present edition),
 - is approved by all parties interested in adherence with IPMVP, and
 - is consistent with the Principles of *M&V* shown in Chapter 3.
3. Follow the approved IPMVP adherent *M&V Plan*.
4. Prepare *M&V* reports containing the information mentioned in the *M&V Reporting* chapter (Chapter 6).

Users wishing to specify the use of IPMVP in an energy performance contract or emission trade may use phrases such as, “The determination of actual energy and monetary savings will follow current best practice, as defined in IPMVP Volume I, EVO 10000 -1:2012.”

Specification may go further to include “The *M&V Plan* shall adhere to IPMVP Volume I, EVO 10000 - 1:2012 and be approved by.....” and may also, if known at the time of contract approval, add, “following IPMVP Option”

CHAPTER 8 OTHER COMMON M&V ISSUES

Beyond the basic framework described in Chapter 4, there are a number of issues which commonly arise regardless of the IPMVP Option chosen. Each of these issues is discussed in this Chapter.

8.1 Applying Energy Prices

Cost *savings*¹⁵ are determined by applying the appropriate price schedule in the following equation:

$$\text{Cost Savings} = C_b - C_r \quad 2)$$

Where:

C_b = Cost of the *baseline energy* plus any *adjustments*¹⁶

C_r = Cost of the *reporting period energy* plus any *adjustments*

Costs should be determined by applying the same price schedule in computing both C_b and C_r . When the conditions of the *reporting period* are used as the basis for reporting energy *savings* (i.e. *avoided energy* use Chapter 4.6.1), the price schedule of the *reporting period* is normally used to compute “avoided cost.”

Examples of the application of energy prices are contained in the examples of Appendix A.

8.1.1 Price Schedules

The price schedule should be obtained from the energy supplier. This price schedule should include all elements that are affected by metered amounts, such as consumption charges, demand charges, transformer credits, power factor, *demand ratchets*, fuel price adjustments, early payment discounts and taxes.

Price schedules may change at points in time different from meter reading dates. Therefore, C_b and C_r in Equation 2) should be computed for periods exactly aligned with price change dates. This alignment may require an estimated allocation of quantities to periods before and after the price change date. The allocation methodology should be the same as that used by the *energy* supplier.

The selected price schedule may be fixed at the date of *ECM* installation, or changed as prices change. (Increasing prices will shorten the *ECM* payback period. Decreasing prices will lengthen the payback period though total *energy* costs will drop when prices drop.) Where a third party has invested in an owner’s *facility*, the price schedule for *savings* reporting is not normally allowed to drop below that which prevailed at the time of commitment to the investment.

¹⁵ See Chapter 9 for the definition of “*savings*.” See also Chapter 4.6 for discussion of the difference between energy *savings* and avoided energy or normalized savings. The same discussion applies to the difference between cost *savings* and cost avoidance or normalized cost savings.

¹⁶ *Adjustments* are the appropriate ones described in Chapter 4.

8.1.2 Marginal Price

An alternative procedure for valuing *savings* involves multiplying *energy* units *saved* by the *marginal price* of *energy*. Be careful to ensure that the *marginal price* is valid for the level of consumption and demand of both the *baseline* and *reporting periods*.

Average, or blended prices, determined by dividing billed cost by billed consumption, are often different from *marginal prices*. In this situation average prices create inaccurate statements of cost *savings* and should not be used.

8.1.3 Fuel Switching and Price Schedule Changes

The Chapter 8.1 general strategy of applying the same price schedule to *baseline* and *reporting period energy* introduces some special considerations when the *ECM* creates a change in fuel type or a change in price schedule between the *baseline* and *reporting periods*. Such situations arise, for example, when an *ECM* includes a change to a lower cost fuel, or shifts the *energy* use pattern such that the facility qualifies for a different price schedule.

In such situations, use the price schedule of the *baseline* commodity to determine C_b in Equation 2). The price schedule of the *reporting-period* commodity should be used in determining C_r . However, both commodity-price schedules would be for the same time period, usually the *reporting period*.

For example, the heating source is changed from electricity to gas, and you intend to use *reporting period* prices. Then C_b would use the *reporting period's* electricity-price schedule for all electricity. C_r would use the *reporting period's* gas-price schedule, for the new gas load, and the *reporting period's* electricity-price schedule for any remaining electricity use.

However, this treatment of an intentional change of price schedule does not apply if the change was not part of the *ECM(s)* being assessed. For example, if the utility changed its price structures for no reason related to the *ECM(s)* being assessed, the Chapter 8.1 general principle of using the same price schedule for C_b and C_r still applies.

8.2 Baseline Adjustments (*Non-Routine*)

Conditions, which vary in a predictable fashion and are significant to *energy* use within the *measurement boundary*, are normally included within the mathematical model used for *routine adjustments*, described in Chapter 4.6. Where unexpected or one-time changes occur in conditions within the *measurement boundary*, which are otherwise static (*static factors*), *non-routine adjustments*, also called *baseline adjustments*, must be made (see also Chapter 4.6).

Non-routine adjustments are needed where a change occurs to equipment or operations within the *measurement boundary* after the *baseline period*. Such change occurs to a *static factor* not to *independent variables*. For example, an *ECM* improved the efficiency of a large number of light fixtures. When more light fixtures were installed, after *ECM* installation, a *non-routine adjustment* was made. The estimated *energy* of the extra fixtures was added to the *baseline energy* so that the *ECM's* true *savings* were still reported.

Values *estimated* for use in IPMVP Option A are usually chosen to eliminate the need for adjustments when changes happen within the *measurement boundary* (see Chapter 4.8.1). Therefore *non-routine adjustments* can be avoided using Option A. For example, a chiller plant's cooling load was *estimated* rather than measured in order to determine Option A *savings* created by a chiller efficiency *ECM*. After retrofit, a *facility* addition increased the actual cooling load within the *measurement boundary*. However since Option A was chosen using a fixed cooling load, reported *savings* are unchanged. The use of Option A avoided the need for a *non-routine adjustment*.

Baseline conditions need to be fully documented in the *M&V Plan* so that changes in *static factors* can be identified and proper *non-routine adjustments* made. It is important to have a

method of tracking and reporting changes in these same *static factors*. This tracking of conditions may be performed by one or more of: the *facility* owner, the agent creating *savings*, or a third party verifier. It should be established in the *M&V Plan* who will track and report each *static factor*.

Where the nature of future changes can be anticipated, methods for making the relevant *non-routine adjustments* should be included in the *M&V Plan*.

Non-routine adjustments are determined from actual or assumed physical changes in equipment or operations (*static factors*). Sometimes it may be difficult to quantify the impact of changes, for example, if they are numerous or not well documented. If the *facility's energy* consumption record is used to quantify the impact of such changes, the impact of the *ECMs* on the *facility's energy* consumption must first be removed by Option B techniques. Option C cannot be used to determine *savings* when the *facility's energy* meter is also used to quantify the impact of changes to *static factors*.

8.3 The Role Of Uncertainty (Accuracy)

The measurement of any physical quantity includes errors because no measurement instrument is 100% accurate. Errors are the differences between observed and true *energy* use. In a *savings*-determination process, errors prevent the exact determination of *savings*. Equation 1) usually involves at least two such measurement errors (*baseline* and *reporting period energy*), and whatever error exists in the computed adjustments. To ensure that the resultant error (uncertainty) is acceptable to the users of a *savings* report, be certain to manage the errors inherent in measurement and analysis when developing and implementing the *M&V Plan*.

Characteristics of a *savings* determination process which should be carefully reviewed to manage accuracy or uncertainty are:

- Instrumentation – measurement equipment errors are due to calibration, inexact measurement, or improper meter selection installation or operation.
- Modeling – the inability to find mathematical forms that fully account for all variations in *energy* use. Modeling errors can be due to inappropriate functional form, inclusion of irrelevant variables, or exclusion of relevant variables.
- Sampling – use of a sample of the full population of items or events to represent the entire population introduces error as a result of: the variation in values within the population, or biased sampling. Sampling¹⁷ may be done in either a physical sense (i.e., only 2% of the lighting fixtures are measured) or a temporal sense (instantaneous measurement only once per hour).
- *Interactive effects* (beyond the *measurement boundary*) that are not fully included in the *savings* computation methodology.
- *Estimation* of parameters using Option A, rather than measurement. You can minimize the variation between the parameter's *estimated value* and its true value through careful review of the *ECM* design, careful estimating of the parameter, and careful *ECM* inspection after installation.

Methods of quantifying, evaluating and reducing some of these uncertainties are discussed in this document's Appendix B. Appendix C lists some references from different regions about applying standard-error-analysis methods to the typical *savings* determination. These quantification tools should only be used to develop the *M&V Plan*, in order to test the inherent uncertainty associated with optional *M&V* program characteristics.

¹⁷ As used in this Protocol, sampling does not refer to rigorous statistical procedures, but to the best practices as addressed in Appendix B-3.

Establish the users' acceptable *savings* accuracy during the *M&V Planning* process. Chapter 8.5 discusses some issues in establishing the correct level of uncertainty for any *ECM* or project. Appendix B-1.2 defines how large *savings* must be, relative to statistical variations in baseline data, for *M&V* reports to be valid.

The accuracy of any measured value is properly expressed as the range within which we expect the true value to fall, with some level of *confidence*. For example, a meter may measure consumption as 5,000 units with a *precision* of ± 100 units, and a *confidence* of 95%. Such statement means that 95% of the readings of the same true value are expected to be between 4,900 and 5,100 units.

In a *savings* determination, it is feasible to quantify many uncertainty factors but usually not all of them. Therefore when planning an *M&V* process, you report both quantifiable uncertainty factors and also qualitative elements of uncertainty. The objective is to recognize and report all uncertainty factors, either qualitatively or quantitatively.

When you describe *precision* within any *savings* report, report the *savings* with no more *significant digits* than the least number of *significant digits* in metered quantities, estimates or constants used in the quantification process. See Chapter 8.12.

8.4 Cost

The cost of determining *savings* depends on many factors such as:

- IPMVP Option selected,
- the number of *ECMs* and the complexity and amount of interaction among them,
- number of energy flows across the *measurement boundary* in Options A, B, or D when applied to a system only,
- level of detail and effort associated with establishing *baseline* conditions needed for the Option selected,
- amount and complexity of the measurement equipment (design, installation, maintenance, calibration, reading, removal),
- sample sizes used for metering representative equipment,
- amount of engineering required to make and support the *estimations* used in Options A or D,
- number and complexity of *independent variables* which are accounted for in mathematical models,
- duration of the *reporting period*,
- accuracy requirements,
- *savings* report requirements,
- process of reviewing or verifying reported *savings*, and
- experience and professional qualifications of the people conducting the *savings* determination.

M&V costs should be appropriate for the size of expected *savings*, the length of the *ECM* payback period, and the report users' interests in accuracy, frequency, and the duration of the reporting process. Often these costs can be shared with other objectives such as real time control, operational feedback, or tenant or departmental sub-billing. Prototype or research projects may bear a larger than normal *M&V* cost, for the sake of accurately establishing the *savings* generated by *ECMs* which will be repeated. However IPMVP is written to provide many

possible ways to document the results of an *ECM* so that users can develop low-cost *M&V* procedures that provide adequate information.

It is difficult to generalize about costs for the different IPMVP Options, since each project will have its own budget. However, *M&V* should incur no more cost than needed to provide adequate certainty and verifiability in the reported *savings*, consistent with the overall budget for the *ECMs*.

**Table 4 Unique Elements
of *M&V* Costs**

Option A	Number of measurement points; complexity of estimation; frequency of <i>reporting period</i> inspections.
Option B	Number of measurement points; length of the <i>reporting period</i> .
Option C	Number of static factors to be tracked during the <i>reporting period</i> ; number of <i>independent variables</i> to be used for <i>routine adjustments</i> .
Option D	Number and complexity of systems simulated; number of field measurements needed to provide input data for the calibrated simulation; skill of professional simulator in achieving calibration.

Table 4 highlights key cost-governing factors unique to each Option, or not listed above.

Commonly, since Option A involves *estimates*, it will involve fewer measurement points and lower cost, providing *estimation* and inspection costs are not unusually high. Option A methods usually have lower cost and higher uncertainty than Option B methods.

Since new measurement equipment is often involved in Options A or B, the cost of maintaining this equipment may make Option C less costly for long *reporting periods*. However, the costs of extra meters for Options A or B may be shared with other objectives of monitoring or cost allocation.

When multiple *ECMs* are installed at one site, it may be less costly to use Options C or D than to isolate and measure multiple *ECMs* with Options A or B.

An Option D simulation model is often time-consuming and costly. However, the model may have other uses such as for designing the *ECMs* themselves or designing a new *facility*.

Expect M&V costs to be highest at the beginning of the reporting period. At this stage in a project, measurement processes are being refined, and accurate performance monitoring helps to optimize *ECM* operation. The cost for each *savings* determination should be in proportion to the expected *savings* and the variation in *savings* (see Chapter 8.5).

A contractor is often responsible for only certain performance indicators. Other indicators may not have to be measured for contractual purposes, though the *facility* owner may still wish to measure all indicators. In this situation, the owner and contractor share the costs of measurement.

8.5 Balancing Uncertainty And Cost

The acceptable level of uncertainty in a *savings* report is related to the cost of decreasing uncertainty to an appropriate level for the expected amount of *savings*. Typically average annual *M&V* costs are less than 10% of the average annual *savings* being assessed. The

quantity of *savings* at stake therefore places a limit on the *M&V* budget, which in turn determines how much uncertainty is acceptable.

For example, consider a project with an expected *savings* of \$100,000 per year and \$5,000/year cost for a basic *M&V* approach with a *precision* no better than $\pm\$25,000$ per year with 90% confidence. To improve the *precision* to $\pm\$7,000$ it may be seen as reasonable to increase *M&V* expenditures up to \$10,000/year (10% of the *savings*), but not \$20,000/year (20%).

The acceptable level of uncertainty in a *savings*-reporting process is often a personal matter, which depends on the report reader's desire for rigor. However reducing uncertainty requires more or better operational data. Enhanced operational data enables fine tuning of *savings* and enhancement of other operational variables. More operational information may also help to size equipment for plant expansions or for the replacement of old equipment.

Enhanced feedback created by *M&V* may also allow higher payments to be made under an *energy performance contract* based on measured data rather than deemed values of *savings*, which must be conservative.

Additional investments for lower uncertainty should not exceed the expected increase in value. This issue is discussed in more detail by Goldberg (1996b).

Of course, not all uncertainties can be quantified (see Chapter 8.3). Therefore both quantitative and qualitative uncertainty statements should be considered when considering *M&V* cost options for each project.

For each project, site and *facility* owner, there is an optimal *M&V Plan*. That optimal *M&V Plan* should include iterative consideration of the sensitivity of the *savings* uncertainty and *M&V* cost to each *M&V* design parameter. Appendix B presents methods of quantifying uncertainty. Appendices B-5.1 and B-5.2 present methods of combining several components of uncertainty and setting uncertainty criteria or objectives.

Not all *ECMs* should expect to achieve the same level of *M&V* uncertainty since uncertainty is proportional to the complexity of the *ECM* and the variations in operations during both the *baseline period* and *reporting period*. For example, Option A methods may allow the *savings* from a simple industrial-plant lighting retrofit to be determined with less uncertainty than the *savings* from a chiller retrofit, since *estimated* lighting parameters may have less uncertainty than *estimated* chiller-plant parameters.

In determining the measurement level and associated costs, the *M&V Plan* should consider the amount of variation in the *energy* use within the *measurement boundary*. For example, indoor lighting may use electricity fairly consistently all year, making it relatively easy to determine *savings*, while heating and cooling loads change seasonally making *savings* identification more difficult. Consider the following general guidelines for balancing cost and uncertainty in an *M&V* process.¹⁸

1. **Low Energy Variation & Low-Value ECM.** Low-value *ECMs* cannot typically afford much *M&V*, based on the 10%-of-*savings* guideline, especially if there is little variation in the measured *energy* data. Such combined situations would tend to favour use of Option A, and short *reporting periods*, for example, in the case of a constant-speed exhaust-fan motor that operates under a constant load according to a well-defined schedule.
2. **High Energy Variation & Low-Value ECM.** Low-value *ECMs* cannot generally afford much *M&V*, as in 1, above. However with a high amount of variation in the *energy* data, the all-parameter measurement techniques of Option B may be needed to achieve the required uncertainty. Sampling techniques may be able to reduce Option B costs. Option C may not be suitable based on the general guidance in Chapter 4.9 that *savings* should exceed 10% of a *facility's* metered use in order to be measurable.

¹⁸ Also see FEMP (2002).

3. **Low Energy Variation & High-Value ECM.** With low variation in *energy* use, the level of uncertainty is often low, so Option A techniques may be most suitable. However, since you expect high *savings*, small improvements in *precision* may have monetary rewards large enough to merit more precise metering and data analysis, if you can keep *M&V* costs appropriate relative to the *savings*. For example, if the *savings* from an *ECM* are \$1,000,000 annually, you may decide to increase the \$5,000 annual *M&V* cost to \$20,000, if it increases *precision* and provides more operational data. Alternatively, a high-value *ECM* may be clearly measurable with Option C. Option C can keep *M&V* costs low, if simple means are used to monitor *static factors* to detect the need for *non-routine adjustments*.
4. **High Energy Variation & High-Value ECM.** This situation allows appropriate uncertainty reduction through extensive data collection and analysis using Options A, B or D. However *savings* are also likely to show in the utility records, so that Option C techniques may be used with careful monitoring of *static factors* to detect the need for *non-routine adjustments*. The *reporting period* may have to span multiple normal cycles of *facility* operation.

8.6 Verification by an Independent Verifier

Where a contractor is hired by a *facility* owner to make and report *energy savings*, the owner may need an independent verifier to review the *savings* reports. This independent verifier should begin by reviewing the *M&V Plan* during its preparation, to ensure that the *savings* reports will satisfy the owner's expectations for uncertainty.

The independent review could also examine *non-routine adjustments*. However full review of *non-routine adjustments* requires good understanding of the *facility*, its operations and *energy* engineering calculation techniques. The *facility* owner should provide summaries of changes in *static factors* so that the verifier can focus on the engineering calculations in the *non-routine adjustments*.

An *energy performance contract* requires that both parties believe that the performance payments are based on valid information. An independent verifier may be helpful to ensure measurement validity and to prevent conflicts. If conflicts arise during the *reporting period*, this independent verifier can help to resolve the conflicts.

Independent verifiers are typically engineering consultants with experience and knowledge in *ECMs*, *M&V* and *energy performance contracting*. Many are members of industry professional societies, or are Certified Measurement and Verification Professionals (CMVP®s).¹⁹

8.7 Data for Emission Trading

Adherence to IPMVP can provide increased confidence in *energy-savings* reports, which also increases confidence in associated reports of emission-reduction commodities.

Combined with the specific *M&V Plan* of each project, IPMVP enhances consistency of reporting and enables validation and verification of energy-saving projects. However to verify an emission-reduction commodity, IPMVP and the project's *M&V Plan* must be used in conjunction with the emission-trading program's specific guidance on converting *energy savings* into equivalent emissions reductions.

Emission trading will be facilitated if the following energy-reporting methods are considered when designing the process for determining the units of *energy saved*.

- Electrical *savings* should be split into peak period and off peak periods, and ozone season and non-ozone season when NO_x or VOC trading is involved. These periods are defined by the relevant emission-trading program.

¹⁹ The CMVP® program is a joint activity of Efficiency Valuation Organization and the Association of Energy Engineers (AEE). It is accessible through EVO's website www.evo-world.org.

- Reductions in purchases from the electrical grid should be divided into those due to load reduction and those due to increased self-generation at the facility.
- The *adjusted baseline* used for computing energy *savings* may need to change to suit the requirements of the particular emission-trading program. For emission-trading purposes, *adjusted baselines* need to consider whether the *ECMs* were 'surplus' or 'additional' to normal behavior. *ECMs* may not be allowed for emission trading if they are 'business as usual' or simply compliance with existing regulations. Baseline rules are defined by the relevant emission-trading program. For example, where equipment minimum-efficiency standards govern the equipment market, these standards set the baseline for determining tradable amounts.
- Segregate energy *savings* by site, if a project spans a power pool's boundary line, or if emission quantities may be outside an air shed of concern.
- Segregate fuel *savings* by fuel or boiler type, if different emission rates apply to each combustion device.

Each emission-trading system usually has its own rules surrounding emission factors to be applied to energy *savings*. For fuel *savings*, default emission rates may be given when no emission-measuring equipment is in place. For electricity *savings*, default values may also be given for the power grid emission rate. Alternatively users may establish their own emission rate for electricity savings, following recognized principles such as those published as part of the Guidelines for Grid-Connected Electricity Projects (WRI 2007).

8.8 Minimum Operating Conditions

An energy-efficiency program should not affect the use of the *facility* to which it is applied, without the agreement of building occupants or industrial process managers. Key user parameters may be: light level, temperature, ventilation rate, compressed air pressure, steam pressure and temperature, water flow rate, production rate, etc.

The *M&V Plan* should record the agreed minimum operating conditions that will be maintained (see Chapter 5).

IPMVP Volume II, Concepts and Practices for Improved Indoor Environmental Quality, suggests methods of monitoring indoor space conditions throughout an energy-efficiency program.

8.9 Weather Data

Where monthly energy measurements are used, weather data should be recorded daily so it can be matched to the actual energy-metering reading dates.

For monthly or daily analysis, government published weather data is usually the most accurate and verifiable. However weather data from government sources may not be available as promptly as site-monitored weather data. If you use on-site weather-monitoring equipment, be sure it is regularly and properly calibrated.

When analyzing the response of energy use to weather in mathematical modeling, daily mean temperature data or *degree days* may be used.

8.10 Minimum Energy Standards

When a certain level of efficiency is required either by law or the facility owner's standard practice, *savings* may be based on the difference between *reporting-period energy* and that minimum standard. In these situations, *baseline-period energy* may be set equal to or less than the applicable minimum *energy* standards.

8.11 Measurement Issues

The proper application of meters for specific applications is a science in itself. Numerous references are available for this purpose (see Chapter 10.2). The EVO web site contains relevant current references on measurement techniques.

Table 5, below, summarizes some key types of meters, and provides comment on *M&V* matters for some of them. This Table should neither be taken as complete nor definitive.

8.11.1 Data Collection Errors and Lost Data

No data collection process is without error. Methodologies for *reporting period* data collection differ in degree of difficulty, and consequently in the amount of erroneous or missing data that may arise. The *M&V Plan* should establish a maximum acceptable rate of data loss and how it will be measured. This level should be part of the overall accuracy consideration. The level of data loss may dramatically affect cost. The *M&V Plan* should also establish a methodology by which missing or erroneous *reporting-period* data will be re-created by interpolation for final analysis. In such cases, *reporting-period* models are needed to interpolate between measured data points so that *savings* can be calculated for each period.

Note that *baseline* data consist of real facts about *energy* and *independent variables* as they existed during the *baseline period*. Therefore *baseline* data problems should not be replaced by modelled data, except when using Option D. Where *baseline* data is missing or inadequate, seek other real data to substitute, or change the *baseline period* so that it contains only real data. The *M&V Plan* should document the source of all *baseline* data.

Table 5
Key Meter
Types –
Part 1

Application	Meter Category	Meter Types	Typical Accuracy	Relative Cost	Best Uses	Special M&V Issues
AC Current (amps)	Current transformer (CT)	Solid torroid or split core transformer	<1%			Not for use where power factor is less than 100% or there is sinewave distortion
AC Voltage (volts)	Voltage leads or 'potential transformer' (PT)	Solid torroid or split core transformer				
AC Electric Power (watts) or AC Energy (watthours)	True rms watt meter or watthour meter	Measure watts (or amps volts and power factor), and watthours. Use digital sampling (IEEE 519-1992) to properly measure distorted waveforms				Necessary for inductive loads (eg motors, ballasts) or circuits with harmonics from components such as a variable speed drive
Runtime (hours)	Measure and record equipment operating periods	Battery operated		Lower cost than watthour recording	Logging of lighting periods	For equipment having a constant power usage rate when on
Temperature (degrees)	Resistance Temperature Detector (RTD)		Reasonable	Low cost	Air and water	Widely used. Take care to compensate for different lead lengths
	Thermo-couple		High	High		Narrow range. Suited to thermal energy metering. Need signal amplifiers

Application	Meter Category	Meter Types	Typical Accuracy	Relative Cost	Best Uses	Special M&V Issues
Humidity (%)						Regular re-calibration required
Liquid Flow (units/sec)	Intrusive	Differential Pressure	1-5% of max			
		Positive Displacement	<1%			
		Turbine, or hot tap insertion turbine	<1%		Clean fluid, straight pipe	
		Vortex Shedding	High			
	Non-intrusive	Ultrasonic	<1%		Straight pipe	Spot flow measurement
		Magnetic		High		
		Bucket & Stop watch		Low	Steam condensate, plumbing outlet fixture	Spot flow measurement
Pressure						
Thermal Energy	Packaged flow and temperature logging and computation	Uses accurate flow and temperature sensors. For steam may need pressure and temperature sensors	<1%	High		Use matched temperature sensors for measuring a temperature difference. Carefully manage all possible sources of error

Table 5
Key
Meter
Types –
Part 2

8.11.2 Use of Control Systems for Data Collection

A computerized control system can provide much of the monitoring necessary for data collection. However, the system's hardware and software must be capable of performing control and data gathering simultaneously, without slowing computer processing, consuming excess communication bandwidth, or overfilling storage.

Some measured parameters may not be useful for control: electric power metering, for example. Trending of small power, lighting and main-feed power consumption may be very useful for high-quality *savings* determination and operational feedback, but useless for real-time control.

Control system software can often perform other functions to assist the tracking of changes to *static factors* during the *reporting period*, such as automatically recording changes in set-points.

Facility staff should be properly trained in this use of the system so they can develop their own trending information for diagnosing system problems, providing the system has the capacity for extra trending. However where a contractor is responsible for some operations controlled by the system, security arrangements should ensure that persons can only access functions for which they are competent and authorized.

The control-system design and monitoring team may have a direct read-only connection into the system via a modem link so that it can easily inspect trend data in the team's office. However possible concerns for virus attacks and computer security should be addressed in this situation.

Control systems can record energy use with their trending capability. However, some systems record "change of value" (COV) events that are not directly used for calculating energy *savings* without tracking time intervals between individual COV events (Claridge et al. 1993, Heinemeier and Akbari 1993). It is possible to tighten COV limits in order to force the trending towards more regular intervals, but this can overload systems that are not designed for such data densities.

Great care should be exercised to:

- Control access and/or changes to the system trend log from which the *energy* data are extracted.
- Develop post-processing routines for changing any control-system COV data into time-series data for performing an analysis.
- Get from the control system supplier:
 - standard traceable calibrations of all sensors it supplies,
 - evidence that proprietary algorithms for counting and/or totaling pulses and units are accurate. (Currently, there are no industry standards for performing this analysis (Sparks et al. 1992)), and
 - commitment that there is adequate processing and storage capacity to handle trending data while supporting the system's control functions.

8.12 Significant Digits

When performing any arithmetic calculation, one must consider the inherent accuracy of the data so the result does not presume greater accuracy than is defensible. For this reason, engineers have adopted a standard of rounding rules that limits the resolution of a result to that which is supported by the data. Therefore, the IPMVP has adopted the following rules to insure all calculations performed under this standard adhere to strict accuracy standards.

Rules for significant digits are derived from the "total derivative" concept from calculus.

Expressed as a function of two variables, the total derivative is, $df(x, y) = \frac{\partial f}{\partial x} \cdot dx + \frac{\partial f}{\partial y} \cdot dy$ 3.1)

If the incremental change, dx & dy , where exchanged for absolute error, Δx & Δy , the following equation results,

$$df(x, y) = \frac{\partial f}{\partial x} \cdot \Delta x + \frac{\partial f}{\partial y} \cdot \Delta y \quad 3.2)$$

From equation 3.2, we can calculate the limits of absolute error. The rules for significant digits agree with equation 3.2 when the absolute error is greater than or equal to ± 1 unit of the smallest significant digit.

To calculate the significant digits of a number, simply count the number of digits ignoring any leading zeros or trailing zeros terminating at the "ones" column (without a decimal point). Any trailing zeros to the right or left of a decimal point are considered significant.

Arithmetic Operation ²⁰	RULE
Addition / Subtraction ²¹ $X + Y$	Round (up or down as appropriate) the result at the right-most decimal place (lowest unit) where all numbers have a common digit. The number of significant digits will be the total of the digits of the result.
Multiplication / Division ²¹ $X \times Y$	The number of significant digits in the result equals the smallest number of significant digits of any one of the input numbers.
Powers X^a	The number of significant digits equals the number of significant digits in the input.

8.12.1 EXAMPLES

Numbers:

- 00123 → 3 significant digits.
- 12300 → 3 significant digits (because it is represented as 1.23×10^4).
- 12300. → 5 significant digits (because it can be represented as 1.2300×10^4).
- 12300.000 → 8 significant digits.
- 12300.012 → 8 significant digits.

²⁰ Additional rules exist for logarithmic and exponential functions that are not included here.

²¹ Mark's Standard Handbook for Mechanical Engineers, 8th Ed., pp. 2.2-2.3

Addition:

$$\begin{array}{r} 0.2056 \\ 2.572 \\ 144.25 \\ + 876.1 \\ \hline \end{array}$$

1,023.1 The number of significant digits is 5.

Multiplication:

- $12.345 \times 0.0369 = 0.456$
- $56.000 \times 0.00785212 = 0.43972$

Powers:

- $3.00^\pi = 31.5$ (3 significant digits in the input generates 3 in the output)

In order to insure consistency and repeatability, all calculations should be carried out by arithmetic operation before applying these rules. For instance, if a motor, running at a constant 32.1 kW ran for 4,564 hrs annually and the utility rate was \$0.0712 per kWh, the cost of the electrical energy is **NOT** ...

$$32.1kW \times 4564hrs = 146,504 \cdot kWh \rightarrow 147,000 \cdot kWh$$

$$147,000 \cdot kWh \times \frac{\$0.0712}{kWh} = \$10,466 \rightarrow \$10,500$$

It is instead correctly calculated by carrying out all the multiplication and division together.

$$32.1kW \times 4564hrs \times \frac{\$0.0712}{kWh} = \$10,431 \rightarrow \$10,400$$

Please also note that the significant digit rules do not mix well together. Carry out all calculations by “arithmetic operation” before proceeding to the next operation type.

8.12.2 SPECIAL CASES

Some numbers are represented with finite significant digits even though they can be treated as exact. Exact numbers have infinite significant digits. An example of an exact number could be a utility rate. If a local power company’s rate was \$0.06 per kWh and Company X used 725,691.0 kWhs one month, the utility bill would be \$43,541.46, not \$40,000 per the multiplication rule above. This is because the utility rate is exact ... it can be represented as ~~\$0.06000~~ per kWh. There is no measurement error associated with utility rates.

Another example includes time variables. If Company X was guaranteed energy savings of \$1.15M per year for 3 years, the total savings would be \$3.45M, not \$3M. Unless expressed as a decimal number, a time variable should be considered as exact.

Care must be taken to recognize these numbers in M&V calculations else the precision of the result may be compromised.

CHAPTER 9 DEFINITIONS

Terms are *italicized* in the text to signify that they have the following meanings:

Adjusted-baseline energy: The *energy* use of the *baseline period*, adjusted to a different set of operating conditions.

Avoided energy use: The reduction in *energy* use that occurred in the *reporting period*, relative to what would have occurred if the *facility* had been equipped and operated as it was in the *baseline period* but under *reporting period* operating conditions. (see Chapter 4.6.1). “Cost avoidance” is the monetary equivalent of “avoided energy use.” Both are commonly called *savings*. *Normalized savings* is another type of *savings*.

Baseline Adjustments: The *non-routine adjustments* (Chapters 4.6 and 8.2) arising during the *reporting period* from changes in any *energy* governing characteristic of the *facility* within the *measurement boundary*, except the named *independent variables* used for *routine adjustments*.

Baseline Energy: The *energy* use occurring during the *baseline period* without adjustments.

Baseline Period: The period of time chosen to represent operation of the *facility* or system before implementation of an *ECM*. This period may be as short as the time required for an instantaneous measurement of a *constant* quantity, or long enough to reflect one full operating cycle of a system or *facility* with variable operations.

Baseline: Pertaining to the *baseline period*.

Confidence Level: The probability that any measured value will fall within a stated range of *precision*. See Appendix B-1.1.

Constant: A term used to describe a physical parameter which does not change during a period of interest. Minor variations may be observed in the parameter while still describing it as constant. The magnitude of variations that are deemed to be ‘minor’ must be reported in the *M&V Plan*.

Commissioning: A process for achieving, verifying and documenting the performance of equipment to meet the operational needs of the *facility* within the capabilities of the design, and to meet the design documentation and the owner’s functional criteria, including preparation of operating personnel.

CV(RMSE): See Appendix B-2.2.2

Coefficient of Variance (cv): See Appendix B-3.1

Cycle: The period of time between the start of successive similar operating modes of a *facility* or piece of equipment whose *energy* use varies in response to operating procedures or *independent variables*. For example the cycle of most buildings is 12 months, since their energy use responds to outdoor weather which varies on an annual basis. Another example is the weekly cycle of an industrial process which operates differently on Sundays than during the rest of the week.

Degree Day: A degree day is measure of the heating or cooling load on a *facility* created by outdoor temperature. When the mean daily outdoor temperature is one degree below a stated reference temperature such as 18°C, for one day, it is defined that there is one heating degree day. If this temperature difference prevailed for ten days there would be ten heating degree days counted for the total period. If the temperature difference were to be 12 degrees for 10 days, 120 heating degree days would be counted. When the ambient temperature is below the reference temperature it is defined that heating degree days are counted. When ambient temperatures are above the reference, cooling degree days are counted. Any reference

temperature may be used for recording degree days, though it is usually chosen to reflect the temperature at which a particular building no longer needs heating or cooling.

Demand: The rate of energy use. Many utilities base a portion of their bills on the highest (or peak) demand they measure during each billing period. Peak demand value are sometimes referred to as simply “demand.” Electrical demand is normally expressed in kilowatts (kW). The sum of monthly billed kW quantities can be expressed in unit of kW-months. See also Demand Ratchet.

Demand Ratchet: A method utilities use to establish the demand for which they invoice when it is different from the demand they meter. Utilities may consider seasonal maximums or minimums, power factor, or contract amounts to set the demand on invoices (called “billing demand”).

Energy: Energy or water use, or demand.

Energy Conservation Measure (ECM): An activity or set of activities designed to increase the *energy* efficiency of a *facility*, system or piece of equipment. ECMs may also conserve *energy* without changing efficiency. Several ECM's may be carried out in a *facility* at one time, each with a different thrust. An ECM may involve one or more of: physical changes to *facility* equipment, revisions to operating and maintenance procedures, software changes, or new means of training or managing users of the space or operations and maintenance staff. An ECM may be applied as a retrofit to an existing system or *facility*, or as a modification to a design before construction of a new system or *facility*.

Energy Performance Contract: A contract between two or more parties where payment is based on achieving specified results, such as reductions in *energy* costs or payback of investment within a stated period.

Energy Services Company (ESCO): A firm which provides services of design and construction of *ECMs* under an *energy performance contract*.

Estimate: A process of determining a parameter used in a *savings* calculation through methods other than measuring it in the *baseline* and *reporting periods*. These methods may range from arbitrary assumptions to engineering estimates derived from manufacturer's rating of equipment performance. Equipment performance tests that are not made in the place where they are used during the *reporting period* are estimates, for purposes of adherence with IPMVP.

Facility: A building or industrial site containing several *energy* using systems. A wing or section of a larger *facility* can be treated as a *facility* of its own if it has meters which separately measure all of its *energy*.

Independent Variable: A parameter that is expected to change regularly and have a measurable impact on the *energy* use of a system or *facility*.

Interactive Effects: *Energy* effects created by an *ECM* but not measured within the *measurement boundary*.

Marginal Price: The cost of one additional unit of a commodity billed under a complex rate schedule.

Mean: See Appendix B-1.3.

Mean Bias Error (MBE): See Appendix B-2.2.2.

Measurement and Verification (M&V): The process of using measurements to reliably determine actual *savings* created within an individual facility by an *energy* management program. *Savings* cannot be directly measured, since they represent the absence of *energy* use. Instead *savings* are determined by comparing measured use before and after implementation of a project, making appropriate adjustments for changes in conditions. See also Chapter 2.

Measurement Boundary: A notional boundary drawn around equipment and/or systems to segregate those which are relevant to *savings* determination from those which are not. All *energy* uses of equipment or systems within the measurement boundary must be measured or estimated, whether the *energy* uses are within the boundary or not. See Chapter 4.4.

Metering: Collection of *energy* data over time at a *facility* through the use of measurement devices.

M&V Plan: The document defined in Chapter 5.

Non-Routine Adjustments: The individually engineered calculations in Equation 1a) of Chapter 4 to account for changes in *static factors* within the *measurement boundary* since the *baseline period*. When non-routine adjustments are applied to the *baseline energy* they are sometimes called just “baseline adjustments.” (See also Chapter 8.2.)

Normalized Savings: The reduction in *energy* use or cost that occurred in the *reporting period*, relative to what would have occurred if the *facility* had been equipped and operated as it was in the *baseline period* but under a normal set of conditions. These normal conditions may be a long term average, or those of any other chosen period of time, other than the *reporting period*. Normal conditions may also be set as those prevailing during the *baseline period*, especially if they were used as the basis for predicting *savings*. (See Chapter 4.6.2) If conditions are those of the *reporting period*, the term *avoided energy use* (see Chapter 4.6.1), or just *savings*, is used instead of normalized savings.

Operational Verification: Verification that the ECMs are installed and operating properly and have the potential to generate savings. *Operational verification* may involve inspections, functional performance testing, and/or data trending with analysis.

Precision: The amount by which a measured value is expected to deviate from the true value. Precision is expressed as a “±” tolerance. Any precision statement about a measured value should include a *confidence* statement. For example a meter’s precision may be rated by the meter manufacturer as ±10% with a 95% confidence level. See Appendices B-1.1 and B-1.2 for definitions of *Absolute Precision* and *Relative Precision*.

Probable Error: See Appendix B-5.

Proxy: A measured parameter substituted in place of direct measurement of an *energy* parameter, where a relationship between the two has been proven on site. For example, if a relationship has been proven between the output signal from a variable speed drive controller and the power requirements of the controlled fan, this output signal is a proxy for fan power.

R Squared (R^2): See Appendix B-2.2.1.

Regression Analysis: A mathematical technique that extracts parameters from a set of data to describe the correlation of measured *independent variables* and dependent variables (usually *energy* data). See Appendix B-2.

Reporting Period: The period of time following implementation of an *ECM* when *savings* reports adhere to IPMVP. This period may be as short as the time for an instantaneous measurement of a constant quantity; long enough to reflect all normal operating modes of a system or *facility* with variable operations; the length of the financial payback period for an investment; the duration of a performance measurement period under an *energy performance contract*; or indefinite.

Routine Adjustments: The calculations in Equation 1a) of Chapter 4 made by a formula shown in the *M&V Plan* to account for changes in selected *independent variables* within the *measurement boundary* since the *baseline period*.

Savings: The reduction in *energy* use or cost. Physical savings may be expressed as *avoided energy use* or *normalized savings* (see Chapter 4.6.1 and 4.6.2, respectively). Monetary savings may be expressed analogously as “cost avoidance” or “normalized cost savings” (see Chapter

8.1). Savings, as used in IPMVP, are **not** the simple difference between baseline and reporting period utility bills or metered quantities. See Chapter 4.1 for elaboration on this point.

Significant Digits: Non-zero digits, and zeroes having non-zero digits to their left. Note that whole numbers (numbers displaying no decimal point) have an unlimited number of significant digits. Whole numbers ending in zero have an unclear number of significant digits. (See also Chapter 8.12.) Note also that when adding numbers, the significant digits rule is replaced by a rule on the number of digits after the decimal place. The number of such digits in any sum should match that of the number with the fewest such digits.

Simulation Model: An assembly of algorithms that calculates *energy* use for a facility based on engineering equations and user-defined parameters.

Standard Deviation: See Appendix B-1.3.

Standard Error: See Appendix B-1.3.

Standard Error of the Coefficient: See Appendix B-2.2.3.

Standard Error of the Estimate: See Appendix B-2.2.2.

Static Factors: Those characteristics of a *facility* which affect *energy* use, within the chosen *measurement boundary*, but which are not used as the basis for any *routine adjustments*. These characteristics include fixed, environmental, operational and maintenance characteristics. They may be constant or varying. (See particularly Chapters 4.6 and 8.2.)

t-statistic: See Appendix B-2.2.3.

Variance: See Appendix B-1.3.

Verification: The process of examining a report prepared by others to comment on its suitability for the intended purpose.

CHAPTER 10 REFERENCES

NOTE: The following references are meant to provide the reader with sources of additional information. These sources consist of publications, textbooks and reports from government agencies, universities, professional organizations and other recognized authorities. For the most part, care has been taken to cite the publication, publisher or source where the document can be obtained.

1. Akbari, H., Heinemeier, K.E., LeConiac, P. and Flora, D.L. 1988. "An Algorithm to Disaggregate Commercial Whole-Facility Hourly Electrical Load Into End Uses", Proceedings of the ACEEE 1988 Summer Study on Energy Efficiency in Buildings, Vol. 10, pp. 10.14-10.26.
2. ASHRAE Guideline 1-1996. The HVAC Commissioning Process. American Society of Heating, Ventilating, and Air Conditioning Engineers, Atlanta, Georgia.
3. ASHRAE Guideline 14-2002, Measurement of Energy and Demand Savings. American Society of Heating, Ventilating, and Air Conditioning Engineers, Atlanta, Georgia.
4. ASHRAE. 1989. An Annotated Guide to Models and Algorithms for Energy Calculations Relating to HVAC Equipment, American Society of Heating, Ventilating, and Air Conditioning Engineers, Atlanta, Georgia.
5. ASHRAE 2005. Handbook: Fundamentals, Chapter 32 - "Energy Estimating and Modeling Methods", Atlanta, Georgia.
6. ASTM 1992. Standard Test Method for Determining Air Leakage Rate by Fan Pressurization, American Society for Testing Materials, Philadelphia, Pennsylvania.
7. Baker, D. and Hurley, W. 1984. "On-Site Calibration of Flow Metering Systems Installed in Buildings", NBS Building Science Series Report No. 159, January.
8. Benedict, R. 1984. Fundamentals of Temperature, Pressure and Flow Measurement. John Wiley and Sons, New York, New York.
9. Benton, C., Chace, J., Huizenga, C., Hyderman, M. and Marcial, R. 1996. "Taking A Building's Vital Signs: A Lending Library of Handheld Instruments", Proceedings of the ACEEE 1996 Summer Study on Energy Efficiency in Buildings, Vol. 4, pp. 4.11-4.21.
10. Bourdouxhe, J.P., Grodent, M., LeBrun, J. 1995. "HVAC01 Toolkit: A Toolkit for 20 Primary HVAC System Energy System Energy Calculations", Final report submitted to ASHRAE.
11. Bourdouxhe, J.P., Grodent, M., LeBrun, J. 1994a. "Toolkit for Primary HVAC System Energy Calculation - Part 1: Boiler Model", ASHRAE Transactions, Vol. 100, Pt. 2.
12. Bourdouxhe, J.P., Grodent, M., LeBrun, J. 1994b. "Toolkit for Primary HVAC System Energy Calculation - Part 2: Reciprocating Chiller Models", ASHRAE Transactions, Vol. 100, Pt. 2.
13. Bou Saada, T.E. and Haberl, J.S. 1995a. "A Weather-Daytyping Procedure for Disaggregating Hourly End-Use Loads in an Electrically Heated and Cooled Building from Whole-facility Hourly Data", 30th Intersociety Energy Conversion Energy Conference, July 30-August 4.
14. Bou Saada, T.E. and Haberl, J.S. 1995b. "An Improved Procedure for Developing Calibrated Hourly Simulated Models", Proceedings of Building Simulation, 1995: pp. 475-484.
15. Bou Saada, T.E., Haberl, J., Vajda, J. and Harris, L. 1996. "Total Utility Savings From the 37,000 Fixture Lighting Retrofit to the USDOE Forrestal Building", Proceedings of the 1996 ACEEE Summer Study, August.

16. Brandemuehl, M. 1993. HVAC02: Toolkit: Algorithms and Subroutines for Secondary HVAC Systems Energy Calculations, American Society of Heating, Ventilating, and Air Conditioning Engineers, Atlanta, Georgia.
17. Bryant, J. and O'Neal, D. 1992. "Calibration of Relative Humidity Transducers for use in the Texas LoanSTAR Program", Proceedings of the 1992 Hot and Humid Conference, Texas A&M University, Energy Systems Laboratory Report No. ESL-PA-92/02-15.
18. Claridge, D., Haberl, J., Bryant, J., Poyner, B. and McBride, J. 1993. "Use of Energy Management and Control Systems for Performance Monitoring of Retrofit Projects", Final Summary Report, USDOE Grant #DE-FG01- 90CE21003, Submitted to the USDOE Office of Conservation and Energy, Energy Systems Laboratory Report ESL-TR-91/09/02, Texas A&M University, March.
19. Claridge, D., Haberl, J., Liu, M., Houcek, J. and Aather, A. 1994. "Can You Achieve 150% of Predicted Retrofit Savings? Is it Time for Recommissioning?", Proceedings of the 1994 ACEEE Summer Study, pp. 5.73- 5.88, August.
20. Claridge, D., Haberl, J., Liu, M. and Athar, A. 1996. "Implementation of Continuous Commissioning in the Texas LoanSTAR Program: Can you Achieve 150% of Estimated Retrofit Savings: Revisited", Proceedings of the 1996 ACEEE Summery Study, August.
21. Cortina, V. (ed.) 1988. "Precision Humidity Analysis", EG&G Environmental Equipment, 151 Bear Hill Road, Waltham, Massachusetts, (IR sensors).
22. Doebelin, E. 1990. Measurement Systems. McGraw-Hill, New York, New York, ISBN 0-07-017338-9.
23. EEI 1981. Handbook for Electricity Metering, Edison Electric Institute, Washington, D.C., ISBN-0-931032-11-3.
24. EPRI 1993. "Fundamental Equations for Residential and Commercial End- Uses" (Rep. #EPRI TR-100984 V2). Palo Alto, California: Electric Power Research Institute.
25. Fels, M. (ed.)1986. "Special Issue Devoted to Measuring Energy Savings, The Princeton Scorekeeping Method (PRISM)", Energy and Buildings, Vol. 9, Nos. 1 and 2.
26. Fels, M., Kisson, K., Marean, M.A. and Reynolds, C. 1995. "Advanced PRISM User's Guide", Center for Energy and Environmental Studies Report, Princeton University, Princeton, New Jersey, January.
27. FEMP – Federal Energy Management Program of the U.S. Department of Energy, 2000. "M&V Guidelines: Measurement and Verification for Federal Energy Projects Version 2.2"
28. FEMP – Federal Energy Management Program of the U.S. Department of Energy, 2002. "Detailed Guidelines for FEMP M&V Option A"
29. Goldberg, M.L. 1996a. "The Value of Improved Measurements: Facing the Monsters That Won't Annihilate Each Other", Energy Services Journal, 2(1):43- 56.
30. Goldberg, M.L. 1996b. "Reasonable Doubts: Monitoring and Verification for Performance Contracting", Proceedings of the ACEEE 1996 Summer Study on Energy Efficiency in Buildings, 4.133-4.143 Washington, D.C.: American Council for an Energy-Efficient Economy.
31. Haberl, J., Bronson, D. and O'Neal, D. 1995. "Impact of Using Measured Weather Data vs. TMY Weather Data in a DOE-2 Simulation", ASHRAE Transactions, V. 105, Pt. 2, June.
32. Haberl, J., Reddy, A., Claridge, D., Turner, D., O'Neal, D. and Heffington, W. 1996. "Measuring Energy-Savings Retrofits: Experiences from the Texas LoanSTAR Program", Oak Ridge National Laboratory Report No. ORNL/Sub/93-SP090/1, February.

33. Haberl, J., Turner, W.D., Finstad, C., Scott, F. and Bryant, J. 1992. "Calibration of Flowmeters for use in HVAC Systems Monitoring", Proceedings of the 1992 ASME/JSES/KSES International Solar Energy Conference.
34. Hadley, D.L. and Tomich, S.D. 1986. "Multivariate Statistical Assessment or Meteorological Influences in Residence Space Heating", Proceedings of the ACEEE 1986 Summer Study on Energy Efficiency in Buildings, Vol. 9, pp. 9.132-9.145.
35. Harding, J. (ed). 1982. "Recent Advances in Chilled Mirror Hygrometry", General Eastern Corporation Technical Bulletin, 50 Hunt Street, Watertown, Massachusetts.
36. Heinemeier, K. and Akbari, H. 1993. "Energy Management and Control Systems and Their Use for Performance Monitoring in the LoanSTAR Program", Lawrence Berkeley National Laboratory Report No. LBL-33114- UC-350, June, (prepared for the Texas State Energy Conservation Office).
37. Houcek, J., Liu, M., Claridge, D., Haberl, J., Katipamula, S. and Abbas, M. 1993. "Potential Operation and Maintenance (O&M) Savings at the State Capitol Complex", Energy Systems Lab Technical Report No. ESL-TR-93/01- 07, Texas A&M University, College Station, Texas.
38. Huang, P. 1991. "Humidity Measurements and Calibration Standards", ASHRAE Transactions, Vol. 97, p.3521.
39. Hurley, C.W. and Schooley, J.F. 1984. "Calibration of Temperature Measurement Systems Installed in Buildings", N.B.S. Building Science Series Report No. 153, January.
40. Hurley, W. 1985. "Measurement of Temperature, Humidity, and Fluid Flow", Field Data Acquisition for Building and Equipment Energy Use Monitoring, ORNL Publication No. CONF-8510218, March.
41. Hyland, R.W. and Hurley, C.W. 1983. "General Guidelines for the On-Site Calibration of Humidity and Moisture Control Systems in Buildings", N.B.S. Building Science Series 157, September.
42. IPCC 1995. Impacts, Adaptations and Mitigation of Climate Change: Scientific-Technical Analyses. Contribution of Working Group II to the Second Assessment Report of the Intergovernmental Panel on Climate Change, Geneva, Switzerland. pp 64.
43. Katipamula, S. 1996. "The Great Energy Predictor Shootout II: Modeling Energy Use in Large Commercial Buildings", ASHRAE Transactions, Vol. 102, Pt 2.
44. Katipamula, S. and Haberl, J. 1991. "A Methodology to Identify Diurnal Load Shapes for Non-Weather-Dependent Electric End-Uses", Proceedings of the 1991 ASME-JSES International Solar Energy Conference, ASME, New York, New York, pp. 457-467, March.
45. Kats, G., Kumar, S., and Rosenfeld, A. 1999. "The Role for an International Measurement & Verification Standard in Reducing Pollution", Proceedings of the ECEEE 1999 Summer Study, Vol. 1, Panel 1.
46. Kats, G., Rosenfeld, A., and McGaraghan, S. 1997. "Energy Efficiency as A Commodity: The Emergence of an Efficiency Secondary Market for Savings in Commercial Buildings", Proceedings of the ECEEE 1997 Summer Study, Vol. I, Panel 2.
47. Kissock, K., Claridge, D., Haberl, J. and Reddy, A. 1992. "Measuring Retrofit Savings For the Texas LoanSTAR Program: Preliminary Methodology and Results", Solar Engineering, 1992: Proceedings of the ASME-JSES-SSME International Solar Energy Conference, Maui, Hawaii, April.
48. Kissock, K., Wu, X., Sparks, R., Claridge, D., Mahoney, J. and Haberl, J. 1994. "EModel Version, 1.4d", Energy Systems Laboratory ESL-SW-94/12-01, Texas Engineering Experiment Station, Texas A&M University System, December.

49. Knebel, D.E. 1983. "Simplified Energy Analysis Using the Modified Bin Method", ASHRAE, Atlanta, Georgia, ISBN 0-910110-39-5.
50. Kulwicki, B. 1991. "Humidity Sensors", Journal of the American Ceramic Society, Vol. 74, pp. 697-707.
51. Landman, D. and Haberl, J. 1996a. "Monthly Variable-Based Degree Day Template: A Spreadsheet Procedure for Calculating 3-parameter Change-point Model for Residential or Commercial Buildings", Energy Systems Laboratory Report No. ESL-TR-96/09-02.
52. Landman, D. and Haberl, J. 1996b. "A Study of Diagnostic Pre-Screening Methods for Analyzing Energy Use of K-12 Public Schools", Energy Systems Laboratory Report No. ESL-TR-96/11-01, November.
53. Leider, M. 1990. A Solid State Amperometric Humidity Sensor, Journal of Applied Electrochemistry, Chapman and Hill: Vol. 20, pp. 964-8.
54. Liptak, B. 1995. Instrument Engineers' Handbook, 3rd Edition: Process Measurement and Analysis. Chilton Book Company, Radnor, Pennsylvania, ISBN 0-8019-8197-2.
55. Miller, R. 1989. Flow Measurement Handbook, McGraw Hill Publishing Company, New York, New York, ISBN 0-07-042046-7.
56. Morrissey, C.J. 1990. "Acoustic Humidity Sensor", NASA Tech Brief. Vol. 14, No. 19, April, (acoustic).
- 56a. ORNL (1999) "A Practical Guide for Commissioning Existing Buildings" Prepared by Portland Energy Coservation Inc., for Oak Ridge National Laboratory (ORNL/TM-1999/34) Available through <http://eber.ed.ornl.gov/commercialproducts/retrocx.htm>
57. Rabl, A. 1988. "Parameter Estimation in Buildings: Methods for Dynamic Analysis of Measured Energy Use", Journal of Solar Energy Engineering, Vol. 110, pp. 52-66.
58. Rabl, A. and Riahle, A. 1992. "Energy Signature Model for Commercial Buildings: Test With Measured Data and Interpretation", Energy and Buildings, Vol. 19, pp.143-154.
59. Ramboz, J.D. and McAuliff, R.C. 1983. "A Calibration Service for Wattmeters and Watt-Hour Meters", N.B.S. Technical Note 1179.
60. Reddy, T. and Claridge, D. 2000. "Uncertainty of "Measured" Energy Savings From Statistical Baseline Models," ASHRAE HVAC&R Research, Vol 6, No 1, January 2000.
61. Reynolds, C. and Fels, M. 1988. "Reliability Criteria for Weather Adjustment of Energy Billing Data", Proceedings of ACEEE 1988 Summer Study on Energy Efficiency in Buildings, Vol. 10, pp.10.237-10.241.
62. Robinson, J., Bryant, J., Haberl, J. and Turner, D. 1992. "Calibration of Tangential Paddlewheel Insertion Flowmeters", Proceedings of the 1992 Hot and Humid Conference, Texas A&M University, Energy Systems Laboratory Report No. ESL-PA-92/02-09.
63. Ross, I.J. and White, G.M. 1990. "Humidity", Instrumentation and Measurement for Environmental Sciences: Transactions of the ASAE, 2nd ed., p. 8-01.
64. Ruch, D. and Claridge, D. 1991. "A Four Parameter Change-Point Model for Predicting Energy Consumption in Commercial Buildings", Proceedings of the ASME-JSES-JSME.
65. SEL 1996. TRNSYS Version 14.2, and Engineering Equation Solver (EES). Solar Energy Laboratory, Mechanical Engineering Department, University of Wisconsin, Madison, Wisconsin.
66. Soebarto, V. 1996. "Development of a Calibration Methodology for Hourly Building Energy Simulation Models Using Disaggregated Energy Use Data From Existing Buildings", Ph.D. Dissertation, Department of Architecture, Texas A&M University, August.

67. Sparks, R., Haberl, J., Bhattacharyya, S., Rayaprolu, M., Wang, J. and Vadlamani, S. 1992. "Testing of Data Acquisition Systems for Use in Monitoring Building Energy Conservation Systems", Proceedings of the Eighth Symposium on Improving Building Systems in Hot and Humid Climates, Dallas, Texas, pp.197-204, May.
68. Vine, E. and Sathaye, J. 1999. "Guidelines for the Monitoring, Evaluation, Reporting, Verification, and Certification of Energy-Efficiency Projects for Climate-Change Mitigation", LBNL Report # 41543.
69. Violette, D., Brakken, R., Schon, A. and Greef, J. 1993. "Statistically-Adjusted Engineering Estimate: What Can The Evaluation Analyst Do About The Engineering Side Of The Analysis?" Proceedings of the 1993 Energy Program Evaluation Conference, Chicago, Illinois.
70. Wiesman, S. (ed.) 1989. Measuring Humidity in Test Chambers, General Eastern Corporation, 50 Hunt Street, Watertown, Massachusetts.
71. Wise, J.A. 1976. "Liquid-In-Glass Thermometry", N.B.S. Monograph 150, January.
72. Wise, J.A. and Soulen, R.J. 1986. "Thermometer Calibration: A Model for State Calibration Laboratories", N.B.S. Monograph 174, January.
73. WRI (2007) Guidelines for Grid-Connected Electricity Projects of the GHG Protocol for Project Accounting, planned for 2007 publication by the World Resources Institute (WRI) and the World Business Council for Sustainable Development (WBCSD), available at www.ghgprotocol.org.

10.1 Resource Organizations

The following US organizations provide useful and relevant information. EVO attempts to maintain on its website (www.evo-world.org) an up to date list of the following, and all other web links in this document:

1. Air Conditioning and Refrigeration Center, Mechanical Engineering, University of Illinois. TEL: 217-333-3115, <http://acrc.me.uiuc.edu>.
2. American Council for an Energy Efficient Economy (ACEEE), Washington, D.C. TEL: 202-429-8873, <http://www.aceee.org>.
3. American Society of Heating, Refrigerating, and Air Conditioning Engineers (ASHRAE), Atlanta, Georgia. TEL: 404-636-8400, <http://www.ashrae.org>.
4. American Society of Mechanical Engineers (ASME), New Jersey. TEL: 800-843-2763. <http://www.asme.org>.
5. Association of Energy Engineers (AEE), Lilburn, GA. TEL: 404-925-9558, <http://www.aeecenter.org>.
6. Boiler Efficiency Institute, Department of Mechanical Engineering, Auburn University, Alabama. TEL: 334/821-3095, <http://www.boilerinstitute.com>.
7. Center for Energy and Environmental Studies (CEES), Princeton University, New Jersey. TEL: 609-452-5445, <http://www.princeton.edu/~cees>.
8. Edison Electric Institute (EEI). Washington, DC. TEL: 202-508-5000, <http://www.eei.org/resources/pubcat>.
9. Energy Systems Laboratory, College Station, Texas. TEL: 979-845-9213, <http://www-esl.tamu.edu>.
10. Florida Solar Energy Center, Cape Canaveral, Florida. TEL: (407) 638- 1000, <http://www.fsec.ucf.edu>.

11. IESNA Publications, New York, New York. TEL: 212-248-5000, <http://www.iesna.org>.
12. Lawrence Berkeley National Laboratory (LBNL), Berkeley CA. TEL: 510- 486-6156, Email: EETDinfo@lbl.gov, <http://eetd.lbl.gov>.
13. National Association of Energy Service Companies (NAESCO), Washington, D.C. TEL: 202-822-0950, <http://www.naesco.org>.
14. Energy Information Administration (EIA), Department of Energy, Washington, D.C., TEL: 202-586-8800, <http://www.eia.doe.gov>.
15. National Renewable Energy Laboratory (NREL), Boulder, Colorado, TEL: (303) 275-3000, <http://www.nrel.gov>.
16. National Technical Information Service (NTIS), U.S. Department of Commerce (This is repository for all publications by the Federal labs and contractors), Springfield Virginia. TEL: 703-605-6000, <http://www.ntis.gov>.
17. Oak Ridge National Laboratory (ORNL), Oak Ridge, Tennessee, Tel: (865) 574-5206, <http://www.ornl.gov/ORNL/BTC>.
18. Pacific Northwest National Laboratory (PNNL), Richland, Washington, Tel: (509) 372-4217, <http://www.pnl.gov/buildings/>.

10.2 Measurement References

ASHRAE (2002) Annex A contains useful information on sensors, calibration techniques, lab standards for measurement, and test methods for chillers, fans, pumps, motors, boilers, furnaces, thermal storage and air handling systems. It also contains useful error and cost considerations, though the cost information is dated because the research that produced the data was done in 1994.

Standards under European Directive 2004/22/EC relative to measurement instruments are:

EN 1359:1998 Gas meters - Diaphragm gas meters
EN 1359:1998/A1:2006
EN 1434-1:2007 Heat meters - Part 1: General requirements
EN 1434-2:2007 Heat meters - Part 2: Constructional requirements
EN 1434-4:2007 Heat meters - Part 4: Pattern approval tests
EN 1434-5:2007 Heat meters - Part 5: Initial verification tests
EN 12261:2002 Gas meters - Turbine gas meters
EN 12261:2002/A1:2006
EN 12405-1:2005 Gas meters - Conversion devices - Part 1: Volume conversion
EN 12405-1:2005/A1:2006
EN 12480:2002 Gas meters - Rotary displacement gas meters
EN 12480:2002/A1:2006
EN 14154-1:2005+A1:2007 Water meters - Part 1: General requirements

EN 14154-2:2005+A1:2007 Water meters - Part 2: Installation and conditions of use
EN 14154-3:2005+A1:2007 Water meters - Part 3: Test methods and equipment
EN 14236:2007 Ultrasonic domestic gas meters
EN 50470-1:2006 Electricity metering equipment (a.c.) - Part 1: General requirements, tests and test conditions - Metering equipment (class indexes A, B and C)
EN 50470-2:2006 Electricity metering equipment (a.c.) - Part 2: Particular requirements - Electromechanical meters for active energy (class indexes A and B)
EN 50470-3:2006 Electricity metering equipment (a.c.) - Part 3: Particular requirements - Static meters for active energy (class indexes A, B and C)

Other European and International standards for measurements and interpretation of data are:

EN ISO 4259 Petroleum products - Determination and application of precision data in relation to methods of test
EN 24185 Measurement of liquid flow in closed conduits - Weighing method (ISO 4185:1980)
EN 29104 Measurement of fluid flow in closed conduits -- Methods of evaluating the performance of electromagnetic flow-meters for liquids
EN ISO 5167 Measurement of fluid flow by means of pressure differential devices - Part 1: Orifice plates, nozzles and Venturi tubes inserted in circular cross-section conduits running full
EN ISO 6817 Measurement of conductive liquid flow in closed conduits - Methods using electromagnetic flow-meters (ISO 6817:1992)
EN ISO 9300 Measurement of gas flow by means of critical flow Venturi nozzles
EURACHEM Quantifying Uncertainty in Analytical Measurement
EUROLAB Technical Report "Measurement Uncertainty – a collection for beginners"
ISO 11453 Statistical interpretation of data - Tests and confidence intervals relating to proportions (1996)
ISO 16269-7 Statistical interpretation of data - Part 7: Median - Estimation and confidence interval (2001)
ISO 3534 Statistics - Vocabulary and symbols
ISO 5479 Statistical interpretation of data - Tests for departure from the normal distribution (1997)
ISO 5725 Accuracy (trueness and precision) of measurement method and results
ISO/TR 5168 Measurement of fluid flow - Evaluation of uncertainties
ISO/TR 7066-1 Assessment of uncertainty in calibration and use of flow measurement devices - Part 1: Linear calibration relationships

See also Annex C for specific measurement standards in various regions of the world.

10.3 Calibration References

Calibration references in Chapter 10's list of US publications above include: ASTM (1992), Baker and Hurley (1984), Benedict (1984), Bryant and O'Neal (1992), Cortina (1988), Doebelin (1990), EEL (1981), Haberl et al. (1992), Harding (1982), Huang (1991), Hurley and Schooley (1984), Hurley (1985), Hyland and Hurley (1983), Kulwicki (1991), Leider (1990), Liptak (1995), Miller (1989), Morrissey (1990), Ramboz and McAuliff (1983), Robinson et al. (1992), Ross and White (1990), Sparks (1992), Wiesman (1989), Wise (1976), Wise and Soulen (1986).

10.4 European and International Standards Supporting Energy Efficiency In Buildings

Assessment of energy performance of buildings on the basis of measured energy use:

- EN15603
- EN 15251
- CEN CR 1752
- ISO/DIS 16814
- ISO 7730

Definitions and requirements related to energy services:

- EN 15900

Economic performance:

- ISO 15686-5, Part 5
- EN 15459

Overall building:

- PrEN15203 (Assessment of delivered energy used in buildings)
- PrEN15603 (Overall energy use primary energy and CO₂ emission)
- PrEN15232 (Calculation methods for energy efficiency improvements by the application of integrated building automation systems)
- EN15316 series (Method for calculation of system energy requirements and system efficiency for heating and domestic hot water)
- ISO 13790 (Thermal performance of buildings – calculation of energy use for space heating)

Heating and cooling calculation and inspection methods:

- EPBD WI 014
- EN14335 series
- EN14243
- ISO 13790
- ISO 16814
- EN13465

- EN13779
- EN15240
- EN15242

Indoor and outdoor condition calculation and presentation of climatic data:

- ISO 15927-1
- ISO 15927-2
- ISO 15927-4
- ISO 15927-5
- ISO 15927-6

APPENDIX A EXAMPLES

A-1 Introduction

This Appendix presents a variety of project types and discusses the key *M&V* design issues arising from the described situations. Each example shows just one IPMVP adherent *M&V* design, though there are numerous possible designs for any project.

The examples cover 12 different scenarios:

- Pump/motor efficiency improvement (A-2)
- Pump/motor demand shifting (A-2-1)
- Lighting efficiency (A-3)
- Lighting operational control (A-3-1)
- Street lighting efficiency and dimming (A-3-2)
- Compressed air leakage management (A-4)
- Turbine-generator set improvement (A-5)
- Boiler efficiency improvement (A-6)
- Multiple *ECM* with metered *baseline* data (A-7)
- Whole facility energy accounting relative to budget (A-7-1)
- Multiple *ECMs* in a building without energy meters in the *baseline period* (A-8)
- New building designed better than code (A-9)

These examples go into varying levels of depth, in order to highlight different features of common *M&V* approaches. None of them is comprehensive. Readers are referred to the EVO website for more complete *M&V Plans* and sample *savings* reports (www.evo-world.org) accessible to EVO Subscribers. Also IPMVP Volume III contains example *M&V* applications for new buildings and renewable-energy projects.

These examples from around the world use the variety of technical units and currencies in local common use. The following table provides an appreciation of the magnitude of the technical quantities expressed in approximate alternate units.

Multiply:		By:	to get:
Natural gas	m ³	35	ft ³
	mcf	1000	ft ³
Steam	pound	0.45	kg of steam
Oil	liter	0.26	gallon (US)

EVO subscribers are encouraged to submit their own examples for possible inclusion in the website's library (email to: ipmvprev@evo-world.org).

A-2 Pump/Motor Efficiency Improvement – Option A

Situation Ten irrigation pump-sets are distributed around a South African agricultural property to pump water from underground wells. Pump operation is usually continuous during the normal six-month annual dry season, though pumps are turned on and off manually as needed. The local utility offered a partial subsidy to replace the pumps with new high-efficiency pumps and

motors. To make the final payment of the subsidy, the utility required short-term demonstration of avoided energy use in a form that adheres to IPMVP. The owner is interested in replacing his old pumps and reducing energy costs, so he paid for the balance of the installation costs and agreed to provide data to the utility after retrofit.

Factors Affecting the M&V Design Pump electricity metering is by 5 utility owned consumption meters. These meters serve only the 10 pumps. Before implementation of the project it was considered possible that the new pumps might enhance pumping rates at some wells, so that pumping hours could be reduced. The owner and the utility recognize that operating hours and therefore *savings* depend upon the growing conditions and rainfall each year. Neither party has control over these energy-governing variables.

The owner sought the lowest possible cost for gathering and reporting information to the utility. The owner hired a contractor to select and install pumps that met his and the utility's specifications.

Pump flow is constant when operating because there are no restricting valves and well depth is largely unaffected by the pumping.

M&V Plan The *M&V Plan* was jointly developed by the owner and utility, following a model provided by the utility. IPMVP Volume I, EVO 10000 – 1:2012, Option A was selected to minimize *M&V* costs. The agreed Option A method is to negotiate an *estimate* of the annual pump operating hours for a normal year, and multiply that number by measured power reductions.

It was agreed that the installation contractor's measurement equipment would be adequately accurate to measure motor wattage requirements. Before removal, the contractor measured the power draw of each old motor after it had been running for at least 3 hours. The utility company maintained the right to witness these measurements. Since the pumps are constant-flow, average annual operating hours were derived from the billed electricity kWh consumption of the past year divided by the measured kW power draw of the old pump motors. This computation showed that on average the pumps operated for 4,321 hours in the dry year before retrofit. The utility found data revealing that total rainfall during that dry season was 9.0% less than normal. The owner and utility therefore agreed that pump operation during that year was 9.0% longer than normal. They agreed that normal hours would be 91.0% of 4,321, or 3,932²² hours per year.

Results The energy *savings* were determined using IPMVP Option A, Equation 1d) as follows:

Total load of all pumps before retrofit:	132 kW
Total load of all pumps after retrofit:	<u>98.2 kW</u>
Net load reduction:	33.8 kW ²³
Energy <i>savings</i> :	$= 34 \text{ kW} \times 3,932 \text{ hours/year} = 130,000 \text{ kWh/year}^{24}$

The utility company's final payment of its subsidy was based on 130,000 kWh energy *savings*.

²² Note this 3,932 number should be expressed with only 3 significant digits, since 91.0% has only 3 significant digits. It should more correctly be expressed as 3.93×10^3 . However common form is used.

²³ The actual calculated number of 33.8 should be treated as having 2 significant digits. This statement is made because the subtraction that led to the 33.8 should show no more digits to the right of the decimal than the number with the fewest to its right (132 has none, so 34 has none).

²⁴ The products of 34 and 3,932 have only 2 significant digits. Though the result is 133,688, the proper expression of their product is 1.3×10^5 , or 130,000.

Using the same *estimated* operating periods, the owner's *estimated savings* under normal rainfall conditions and at current utility prices were determined to be $34 \text{ kW} \times 3,932 \text{ hr}^{25} \times \text{R}0.2566/\text{kWh} = \text{R}34,000/\text{year}$.²⁶ Utility service and network charges were unchanged.

A-2.1 Pump/Motor Demand Shifting – Option B

Situation The irrigation system described in Appendix A-2 above was also eligible for a substantial utility incentive if the pumps are kept off during the peak periods of 0700-1000 and 1800-2000 on all weekdays that are not public holidays. The owner installed a radio-signal-based control system to remotely and automatically control the pumps to implement this load shifting strategy. The pump control will be reset by the owner annually according to the upcoming year's schedule of public holidays.

Factors Affecting the M&V Design The owner believed that curtailing pumping for a maximum of 25 hours per week (15%) would not be critical to his operation in dry seasons. (He expected fewer breakdowns of the new pumps, so there would be no net impact on his dry-season growth.)

The utility recognizes that the owner decides whether to shut down the pumps based on his own needs. Therefore the utility required adherence to IPMVP Volume I, EVO 10000 – 1:2012, Option B to substantiate each year's performance, before making the incentive payment.

The owner felt that his financial payback period for the control and monitoring equipment was already long. Therefore he does not want to spend a significant part of the incentive on providing the evidence required by the utility.

M&V Plan The utility and owner agreed that continuous recording of a *proxy* variable would give the ongoing evidence that the pumps were off during every peak period all year long. The *proxy* variable is the presence of electricity flow (in excess of the 500mA needed by the control equipment) through any of the 5 electrical feeds to the 10 pumps. Small un-calibrated current sensors and data loggers were clamped on each power line near the 5 meters. The sensors and loggers have a re-chargeable battery-backup power system.

The owner has hired the supplier of the control and monitoring devices to annually read the data, check the clock settings, and give a report to the utility of the dates and times of any operation within any weekday peak periods.

Results For the first year after implementation of the control and monitoring system, the monitoring agent reported to the utility that power was used between 1800 hrs and 2000 hrs on 5 specific weekdays. The utility verified that these days were all public holidays, so there were no operations during the defined peak periods. The demand shift was determined to be 98.2 kW, from the measurement of the new pumps (see Appendix A-2). The annual utility incentive was computed and paid based on this Option B recorded 98.2 kW demand shift.

A-3 Lighting Efficiency – Option A

Situation More efficient light fixtures are installed in place of existing fixtures in a Canadian school, while maintaining light levels. This project was part of a broader program of the school board to hire a contractor, who would design, install and finance many changes in a number of schools. Payments under the contract are based on measured *savings* at the utility prices prevailing at the time of signing the contract. *Savings* are to be demonstrated, according to an IPMVP adherent *M&V Plan*, immediately after commissioning of the retrofit. Since the owner controls operation of the lights, the contract specified that the *M&V Plan* follow IPMVP Volume I,

²⁵ 133,688 is the actual calculated value before significant digit rounding.

²⁶ This amount can be expressed in no more than 2 significant digits, as from the above observations about the minimum number of significant digits. The actual calculated value is R34,103 and should better be expressed as R3.4 x 10⁴, though 34,000 is customary currency format.

EVO 10000 – 1:2012, Option A, using *estimated* operating hours. The *M&V Plan* was to be detailed after contract signing.

Factors Affecting the M&V Design In developing the *M&V Plan* the following were considered:

- All light fixtures are powered by a common 347-volt supply system dedicated to lighting. This situation makes power measurement simple.
- Operation of lights significantly affects heating energy requirements, so the interactive effect needed to be estimated.
- Operation of lights significantly affects mechanical-cooling requirements. However, since very little of the school is mechanically cooled and that space is usually vacant during the warmer weather, cooling *interactive effects* were ignored.
- School-board officials had difficulty accepting an arbitrary assumption of lighting operating periods. They agreed to pay for a carefully instrumented two-month period of logging lighting patterns in one school. This test would substantiate the *estimated* operating hours that would be agreed for all schools.

M&V Plan The *measurement boundary* of this *ECM* was drawn to include the lighting fixtures connected to the 347-volt supply system.

- The heating *interactive effect* was determined by engineering calculations to be a 6.0% increase in boiler-output energy requirements, for the period from November through March. Boiler efficiency in winter was estimated to be 79% under typical winter conditions.
- The *static factors* recorded for the *baseline* included a lighting survey giving a description, location, light level, and count of the number of operating and burned out lamps ballasts and fixtures.
- 30 lighting loggers were placed in randomly chosen classrooms, corridors, locker rooms, and offices and also in the gym and auditorium, for two months. This period included the one-week spring holiday and two legal holidays. Table A-3-1 summarizes the data obtained.

Location	Fraction of Lighting Load	Mean weekly hours	
		School Time	Holiday Time
Locker rooms	5%	106.	22.
Offices	5%	83.	21.
Classrooms	61%	48.	5.
Auditorium	10%	31.	11.
Gymnasium	10%	82.	25.
Corridors	9%	168.	168.

Table A-3-1
Operating Period
Survey

Since classrooms are the largest load, the *relative precision* of the classroom operating period measurements was evaluated before school board officials could agree to *estimated* values. For the 19 classroom loggers, the *standard deviation* among the readings for 6 recorded school weeks was found to be 15 hours per week. With $19 \times 6 = 114$ readings, the *standard error* in the *mean* values was computed to be 1.4 hours per week (Equation B-4). At 95% *confidence*, the value of *t* for a large number of observations is 1.96 (Table B-1). Therefore, using Equation B-9, it was established with 95% *confidence* that the *relative precision* in the measured classroom operating hours is:

$$= \frac{1.96 \times 1.4}{48} = 5.7\%$$

School board officials deemed this measurement *precision* adequate.

Before estimating values for all schools, it was decided to add 6 hours per week to classroom hours because of plans to increase night school classes. Considering that there are 39 school weeks and 13.2 holiday weeks in an average year (with leap years), the *estimated* annual operating hours were agreed to be as follows:

Table A-3-2
Estimated
Operating Hours

Location	Fraction of Lighting Load	<i>Estimated Weekly Hours</i>		<i>Estimated Annual Hours</i>
		39.0 school weeks	13.2 holiday weeks	
Locker rooms	5.00%	106	22	4,420
Offices	5.00%	83	21	3,480
Classrooms	61.00%	54	5	2,170
Auditorium	10.00%	31	11	1,350
Gymnasium	10.00%	82	25	3,530
Corridors	9.00%	168	168	8,770

Since the lighting retrofit was applied uniformly to all fixtures, the load-weighted average *estimated* annual operating hours for this school were determined to be 2,996, or 3,000 rounded to 3 significant digits (a better representation of the result would be 3.00×10^3).

- *Baseline* power measurements were made with a recently calibrated true rms watt meter of the three-phase power draw on the 347-volt lighting circuits. From a thirty-second measurement on the input side of two lighting transformers, it was found that with all fixtures switched on, the total power draw was 288 kW. Seventy lamps (= 3 kW or 1%) were burned out at the time of the test. It was determined that the fraction burned out at the time of this measurement was normal.
- Since lighting loads establish the building electrical peak at a time when all lights are on, electrical demand *savings* will be estimated to be the same as the measured load reduction on the lighting circuits. The utility bills showed a lower demand during the summer holidays, and there was minimal use of the facility during these months. Also considering which other equipment was used during the summer, it was assumed that the July and August lighting-circuit demand is only 50% of the peak measured circuit load.
- The marginal utility prices at the time of contract signing was CDN\$0.063/kWh, CDN\$10.85/kW-month, and CDN\$0.255/m³ of gas.

Results After installation of the *ECM*, the lighting circuit power was re-measured as in the *baseline* test. The power draw was 162 kW with all lights on and none burned out. With the same 1% burnout rate as in the base year, the post-retrofit period maximum power would be 160. kW (=162 x 0.990). Therefore the power reduction is 288. – 160. = 128 kW.

Energy *savings* (using Equation 1d) with no adjustments) are 128 kW x 3.00×10^3 hrs/year = 384,000 kWh/year.

Demand *savings* are 128 kW for 10.0 months and 64 kW for 2.0 months, for a total of 1,410 kW-months.

The value of the electrical *savings estimated* under IPMVP Option A is:

$$(384,000 \text{ kWh} \times \$0.0630) + (1,410 \times \$10.85) = \text{CDN\$}39,500$$

Assuming the lighting *savings* are achieved uniformly over a 10 month period, the typical winter month electrical *savings* are $384,000/10 = 38,400$ kWh/month. The associated boiler load increase is 6.0% of these electrical *savings* for November through March, namely:

$$= 6.0\% \times 38,400 \text{ kWh/mo} \times 5.0 \text{ months} = 12,000 \text{ kWh}$$

Extra boiler input energy is:

$$= 12,000 \text{ kWh} / 79\% = 6.0\% \times 38,400 \text{ kWh/mo} \times 5.0 \text{ months} / 79\% = 15,000 \text{ kWh}$$

equivalent units of fuel input.

The gas being used in the boiler has an energy content of 10.499 kWh/m^3 , so the amount of extra gas is $= 15,000 / 10.499 = 1,400 \text{ m}^3$ gas

The value of the extra gas used in winter is $1,400 \times \$0.255 = \text{CDN}\$360$. Therefore total net *savings* are $\$39,500 - \$360 = \text{CDN}\$39,100$.

A-3-1 Lighting Operational Control – Option A

Situation A knitting mill in southern India typically operates 2 shifts per day. There was a standing order for the supervisors to turn off all lights in each zone at the end of the second shift. There are 70 light switches. Supervisors regularly changed between working on the first and second shifts. They habitually forgot their duty to turn off lights.

The plant manager undertook a project to modify the lighting so that occupancy sensors turned lights on and off. He wanted to document the results to show the supervisors how poorly they had been using the light switches.

Factors Affecting M&V Design None of the production area had windows or skylights. It is neither heated nor cooled. Lighting circuits are integrated with other electrical loads so that lighting use could not be easily isolated from other uses of electricity.

The plant manager did not wish to spend a lot to determine *savings*, but needed a credible statement of the *savings*.

The electricity price for medium sized commercial users is 450 p/kWh.

M&V Plan To minimize M&V costs it was decided to perform *savings* measurements for only a short representative period and use IPMVP Volume I, EVO 10000 – 1:2012, Option A. Since the primary purpose of the retrofit was to control production area lighting hours, a sampling based method was developed to measure the change in operating hours. The lighting power (for use in Equation 1d)) was *estimated* from manufacturers' ratings to be 223 kW.

Lighting loggers were placed randomly around the production area to record the operating hours of randomly chosen lighting zones. The number of loggers was chosen as follows, to obtain an overall *precision* in operating period estimates of $\pm 10\%$, at 90% *confidence*. It was expected that the *mean* operating hours before installation of the occupancy sensors would be 125 hours per week, and that the *standard deviation* in readings would be 25. Therefore the initially estimated *cv* is 0.20 and the necessary number of samples (with *z* of 1.96) is 15 (Equation B-16). Since there are only 70 zones, the finite population adjustment lowers the estimated required number of loggers to 12 (Equation B-17). It was assumed that after installation of occupancy sensors the *cv* will be much lower so the 12 loggers will be adequate.

There are no *interactive effects* of this retrofit on other building loads because the plant is neither heated nor air-conditioned. The reduction in night-time lighting is expected to make the building more thermally comfortable at the beginning of the morning shift.

Results After a one month period, data was gathered from the loggers and the average weekly operating hours computed for the 12 zones. The *mean* value was 115 and the *standard deviation* was 29. Therefore the *cv* was 0.24 ($= 29 / 115$), higher than the expected value and worse than necessary to meet the *precision* requirement. Therefore another month of recording was undertaken. Then the *mean* of the eight weeks of average weekly values was 118, and the

standard deviation was 24 (*cv* = 0.20). This was deemed an adequate measurement of operating hours in the *baseline period*, with no occupancy sensors.

The occupancy sensor controls were installed after the above *baseline* test. Operating hours were again logged in the same locations for a month. The *mean* was found to be 82 hours per week, and the *standard deviation* was 3 hours. In this situation the *cv* is 0.04 and well within the required 0.2, so the one-month readings were accepted. No changes had happened to the way the plant was used or occupied, so there is no need to make any *non-routine adjustment* to the *baseline* data.

The reduction in operating hours was $118 - 82 = 36$ hours per week. *Savings* were computed using Equation 1d) as:

$$223 \text{ kW} \times 36 \text{ hours/week} = 8,000 \text{ or } 8.0 \times 10^3 \text{ kWh/week}$$

With 48 weeks of operation every year, the annual value of the consumption *savings* is:

$$= (8.0 \times 10^3) \times 48 \times 450 / 100 = \text{Rs } 1.7 \text{ million}$$

There are no demand *savings* since the retrofits only affect off peak power use.

Therefore, following IPMVP Option A, it can be stated with 90% *confidence* that the *savings*, in the month after occupancy sensor installation, were Rs17 lakh $\pm 10\%$, given the estimate of installed lighting load.

A-3-2 Street Light Efficiency and Dimming – Option B

Situation A Croatian city's public lighting system was in need of substantial repair and updating. A new lighting system was installed on the same wiring, including high-efficiency fixtures and a dimming system which curtails lighting power by up to 50% in the quietest hours. The lighting is distributed across the city, with 23 metering points. The retrofit included the addition of centralized dimming control. The city retained the current lighting-maintenance contractor to design, install and maintain the system. The city obtained a *savings* performance guarantee from the contractor. The city required the contractor to continuously demonstrate achievement of the guaranteed *savings*.

Factors Affecting The M&V Design The *baseline* light levels were inconsistent because 20% of the fixtures were burned out. The city wished to maintain a more uniform light level. Therefore it upgraded its public lighting maintenance contract to specify that burnouts be no more than 3.0% at any time.

Since dimming is critical to the *savings*, continuous recording of energy use is required. The 23 utility meters measure energy use continually. However these meters cannot provide the rapid operational feedback necessary to avoid significant energy wastage if a dimmer fails or is accidentally changed. Consequently an energy recording capability was added to the central dimming control system, to remotely record energy use in the city's central control station. Beyond simple energy reporting, the system compares actual hourly energy use on each circuit to an expected hourly profile. Variances from this target are used to spot burnouts and failures of the dimming system.

M&V Plan *Baseline* electricity on all 23 utility meters for the past year totaled 1,753,000 kWh, from utility bills. The number and location of all fixtures in the *baseline period* was recorded as part of the *M&V Plan*, along with the operating setpoints of the lighting control system.

Annual energy, recorded on the bills for same accounts will be totaled for determining *savings* using IPMVP Volume I, EVO 10000 – 1:2012, Option B, Equation 1c). The only adjustments that will be made to *baseline* or *reporting period* energy use will be for additions or deletions to the system and for burnouts found to be more than 3% at any time.

A *non-routine adjustment* was made immediately to account for reducing the burnout rate from the *baseline period's* 20% to the target *reporting-period* value of 3.0%. The *baseline-energy* was therefore adjusted to 2,130,000 kWh (= 1,753,000 x 0.970 / 0.800).

The city staff will monitor burnout rates monthly. If the burnout rate is greater than 3.0%, a *non-routine adjustment* will be made to bring *reporting-period* metered data up to the contracted 3.0% burnout rate.

Savings will be reported for the length of the 10-year guarantee period using a single price of 0.600 kuna/kWh.

Results *Savings* were reported without adjustment for the first three years after retrofit because burnout rates remained above 3.0%.

For the fourth year the burnout rate was 5.0% for 7 months. Fourth year *savings* were computed as follows:

Baseline Energy 2,130,000 kWh

Fourth year measured energy = 1,243,000 kWh

The burnout adjustment is =

$$\left(\frac{0.970}{0.950} - 1.000 \right) \times \frac{7.0}{12} \times 1,243,000 = 15,000 \text{ kWh}$$

Adjusted fourth year energy = 1,243,000 + 15,000 = 1,258,000 kWh

Savings (avoided energy) = 2,130,000 – 1,258,000 = 870,000 kWh

Avoided Cost = 870,000 kWh x 0.600 = kn 520,000

A-4 Compressed-Air Leakage Management – Option B

Situation A Brazilian auto manufacturer's plant engineering department estimated that R\$200,000 per year was being lost through compressed-air leakage arising from poor maintenance. The plant engineer persuaded the plant manager that the maintenance department should dedicate one person for two months to repair all leaks. The engineering department agreed to conduct ongoing monitoring of leakage rates and savings, in order to motivate the maintenance staff to regularly check for leakage.

Factors Affecting M&V Design There are very few funds available for any *M&V* activity. Also the engineering department wished any *savings* measurement methodology to have a maximum quantifiable error of ±5% in any reported *savings*, with a *confidence* level of 95%.

The plant operates with 2 shifts per day, 10 per week and 442 per year. When it is operating, it's use of compressed air is steady. Heat from compressors is rejected directly outside compressor rooms without impacting any other plant energy-using systems.

The local electric consumption rate (known as the "green rate") for low-load-factor commercial accounts over 0.5 MW is shown in Table A-4-1.

	Dry Months (May – September)	Wet Months (October – April)
Peak Periods (17:30-20:30 hrs Monday to Friday)	R\$0.957/kWh	R\$0.934/kWh
Off Peak Periods	R\$0.143/kWh	R\$0.129/kWh

**Table A-4-1 Electric
Consumption Prices**

Taxes totaling 42.9% are added to these rates.

It was assumed that the impact on plant electrical demand would be minimal since it is likely that there will be no change in the maximum number of compressors that will function during plant operations.

M&V Plan A full *M&V Plan* is shown on EVO's subscriber website (www.evo-world.org). It uses IPMVP Volume I, EVO 10000 – 1:2012, Option B, for ongoing measurement of *savings* to indicate changes in compressed-air leakage rates. IPMVP Equation 1b) was used to adjust *baseline energy* to *reporting period* conditions. The *M&V Plan* aimed to minimize extra measurement costs so a simple three-phase true-rms wattmeter was added to the electrical supply of the motor-control center feeding all equipment in the compressor room. This *measurement boundary* encompassed 6 compressors, 3 compressed-air driers, and all other minor auxiliary systems in the compressor room. Heat generated within the compressor room is not an *interactive effect* since it does not affect any other energy uses. Plant staff were instructed to read the meter at the end of each shift (i.e. three times a day) whether the plant was operating or not. The meter was installed three months before leak-management activities began.

The *static factors* related to plant design and operations were listed, as a reference for any future possible *non-routine adjustments*. They included the number, capacity and usage patterns of all compressed-air-driven equipment, plant production-line speed, and vehicle models being produced.

The *baseline period* electricity use, for operating and non-operating shifts, were quite different. Also within either kind of shift there were slight variations in energy use. No specific *independent variable* could be identified to account for the variations. It was decided to use the *mean* energy use of each kind of shift in the *baseline period* for determining the *savings*. A criterion was established for determining when sufficient readings had been made of *baseline energy* per shift to meet the target 95/5 uncertainty target for any *savings* report.

Results A full set of *savings* results are shown on the EVO subscriber website. It was found that to meet the 95/5 uncertainty criterion, the variation in shift energy during the *baseline* required readings for a seven-week period before retrofit. The *baseline* values were therefore established as the seven-week average electricity use of operating and non-operating shifts.

It was noted that after the leakage repair activity was completed there was much less variation in the *reporting-period* energy use per shift. Therefore the uncertainty target could be met by monthly *savings* reports.

Energy *savings* were computed as the difference between actual energy use every month and the *adjusted-baseline* energy determined by multiplying the number of actual shifts in the month by the *baseline mean* energy use for each type of shift.

The appropriate price of electricity was applied to the consumption *savings*, assuming that the utility's "peak period" rates only applied to three hours within the second shift. No demand *savings* were calculated.

These measurements continued as part of normal plant operations. The plant-engineering department adjusted the *baseline energy* periodically as *static factors* changed. Operating staff provided shift energy readings and the engineering department reported *savings* every month. Variations from past *savings* patterns became a focus for assessing the maintenance practices related to the compressed-air system.

A-5 Turbine/Generator Set Improvement – Option B

Situation A pulp mill used a steam turbine to generate much of its own electricity. Recent process changes had reduced the available steam for the turbine-generator (TG) unit from its original design level. As a result electricity output and thermal efficiency of the TG unit was reduced. The mill installed a new more efficient rotor designed for the new smaller steam flow. A

measurement process was put in place for assessing the increased electrical output in order to qualify for an electric-utility incentive payment.

Factors Affecting The M&V Design The purpose of the M&V was to report electrical improvements. The mill recognized that extraction of more energy by the turbine left less steam energy for the process, or required more boiler energy to deliver the same steam to the process. These *interactive effects* were not part of this analysis for the electrical utility. The utility incentive was based purely on increased electricity production.

M&V Plan The mill and the utility agreed to use IPMVP Volume I, EVO 10000 – 1:2012, Option B to determine the increase in electricity output for a one year period. Existing plant instrumentation was used to determine the efficiency of the old rotor as shown in Figure A-5.1.

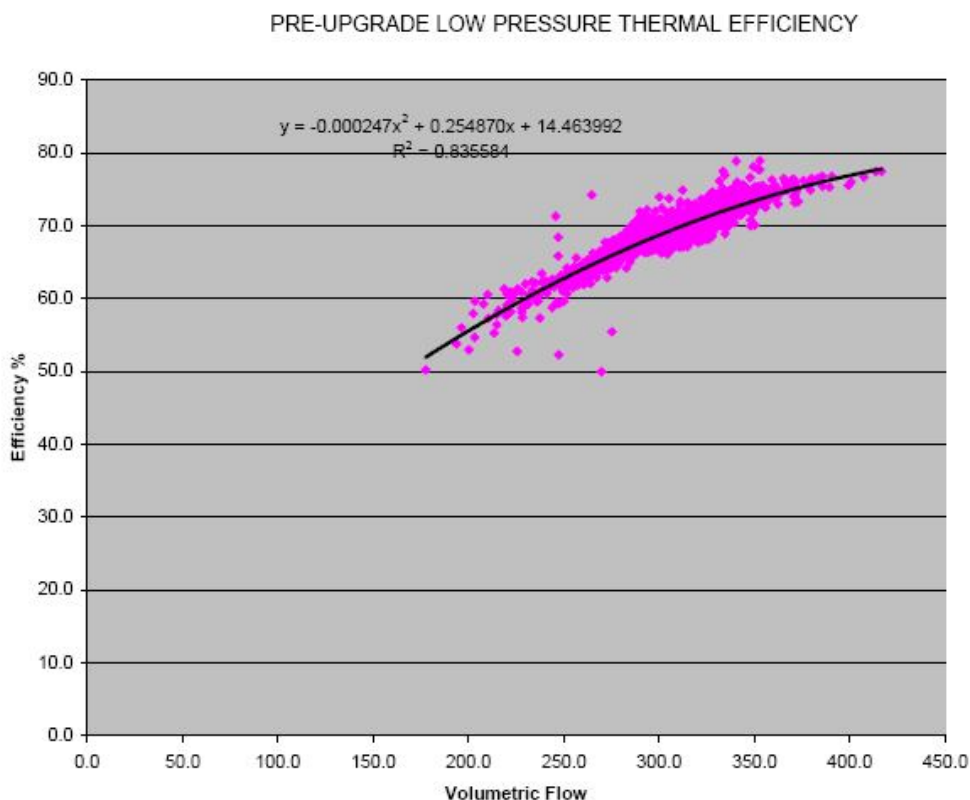


Figure A-5.1
Old Rotor
Performance

The mathematical model describing the *baseline* unit efficiency was found by regression analysis to be:

$$\text{Efficiency (\%)} = (-0.000247 \times \text{flow}^2) + (0.255 \times \text{flow}) + 14.5$$

This efficiency model will be used with the steam conditions of the one-year *reporting period* to determine what the electricity production would have been with the old rotor. Increased electricity production will be reported under *reporting-period* conditions, using IPMVP Equation 1b).

Existing plant meters are regularly calibrated as part of plant maintenance. They were deemed to be suitable for the utility's purpose.

Results For a year after retrofit, the steam conditions every minute were applied to the mathematical model of old rotor efficiency to compute the *adjusted-baseline energy* term used

in IPMVP Equation 1b). This value was compared to actually measured generation for the same period to determine the increase in electrical output.

No changes happened to the TG unit during this year, so *non-routine adjustments* were unnecessary.

A-6 Boiler Efficiency Improvement – Option A

Situation A boiler contractor replaced a German office building's existing boiler with a more efficient boiler. The contractor guaranteed annual oil *savings* of at least €25,000, assuming the loads on the boiler were the same as he measured during the *baseline period*. The owner's purchase order specified that holdback amounts would be paid only after the contractor presented a *savings* report adhering to IPMVP Volume I, EVO 10000 – 1:2012. It was also specified that the owner and contractor would agree to the *M&V Plan* as part of the final design plans for the retrofit.

Factors Affecting M&V Design Numerous building changes were going on at the time of the boiler plant revision, so boiler plant loads were expected to change. The contractor is only responsible for boiler efficiency improvements, not changes in boiler load. The boiler is the only equipment in the building using oil. The price of oil to be used for proof of achieving the performance guarantee was €0.70/liter.

M&V Plan IPMVP Volume I, EVO 10000 – 1:2012, Option A was chosen to isolate the boiler from the changes going on in the rest of the building. The *measurement boundary* was drawn to include only the boiler, measuring fuel use and net thermal energy delivered to the building. This boundary excludes the electricity use of the boiler's burner and blower. Changes to these electrical *interactive effects* were regarded as negligible, and not worth inclusion within the *measurement boundary* or even separate estimation.

The contractor's guarantee was stated relative to the usage of the year before submission of its proposal. During that period, the facility purchased 241,300 liters of heating oil for the boiler. There was a 2,100 liter increase in oil inventory between the beginning and end of that year. Therefore actual consumption was 239,200 liters. The energy load on the boiler will be determined from this oil-use data, once the efficiency of the old boiler is established. IPMVP Equation 1d) will be used with 239,200 liters as the *estimate*. This *estimate* has no error, since most of it²⁷ comes from oil shipment data, which is the reference source with no error.

Boiler efficiency will be the measured parameter in Equation 1d). Efficiency tests were planned for a period of typical winter conditions before removing the old boiler. Winter conditions were chosen so that there was sufficient load to assess efficiency under the full range of boiler loads. A recently calibrated thermal energy meter was installed by the contractor on the boiler supply and return water lines and a calibrated oil meter installed on the fuel supply to the boiler. Both the oil meter and the thermal-energy meter and data logger have manufacturers' rated *precisions* of $\pm 2\%$ for the ranges involved in this project.

Baseline efficiency tests were conducted over three separate one-week periods when daily *mean* ambient temperatures ranged from -5°C to $+5^{\circ}\text{C}$. Identical tests were planned for the first period after commissioning of the new boiler when ambient temperatures are once again in the -5°C to $+5^{\circ}\text{C}$ range, using the same oil and thermal energy meters left in place since the *baseline* efficiency tests. Since the three individual one-week tests are expected to include periods

²⁷ Oil inventory levels are read from an un-calibrated tank gauge of unknown accuracy. Since the magnitude of inventory adjustments are small relative to metered shipments for the year, any error in this inventory term were considered negligible.

representing a range of boiler loads, from low to high, it was agreed that the test results will adequately represent the annual improvement that the owner could expect.

Oil and thermal energy meter readings will be made daily by building maintenance staff through winter months until three valid weeks of testing have been obtained for the old boiler. The same process will be followed for the new boiler. The readings will be logged in the boiler room and open for inspection at any time. The building-automation system measures and records ambient temperature for the valid weeks.

A contract extra of €5,100 was accepted by the owner for the supply, installation and commissioning of the oil and thermal-energy meters and for computing and reporting the *savings*. Consideration was given to requiring demonstration of performance for a whole year. However the contractor pointed out that the extra costs of meter calibration and data analysis would add €3,000 to the fee. The owner decided that a short test period of 3 representative weeks would be adequate. The owner also decided to maintain and calibrate the oil and thermal energy meters himself after the contract, and to annually make his own boiler-efficiency calculations.

Results *Baseline* oil and thermal energy data was collected continuously over a five-week period, until three were found where daily mean ambient temperatures stayed within the specified range -5°C to +5°C. Dividing net thermal energy delivered by oil consumed, the average efficiency readings for the old boiler during the three one-week periods were found to be 65.2%.

After installation and commissioning of the boiler, the three-week *reporting period* was again found with an average ambient temperature between -5°C to +5°C. Boiler efficiency test results averaged 80.6%.

There were no other changes to the boiler plant between the time of the *baseline-period* tests and *reporting-period* tests. Therefore *non-routine adjustments* were not needed.

Using IPMVP Equation 1d), annual *savings* using 239,200 liters as the estimated annual oil use from the baseline period are:

$$\begin{aligned}\text{Oil savings} &= 239,200 \text{ liters} \times (1.000 - 0.652 / 0.806) \\ &= 45,700 \text{ liters}\end{aligned}$$

The value of the *savings* is €0.70 x 45,700 = €31,900.²⁸

These estimated annual *savings* from a short-term test validated that the contractor had met its guaranteed performance.

A-7 Multiple ECM With Metered *Baseline* Data – Option C

Situation An energy efficiency project was implemented in a high school in northern United States. It involved ten *ECMs* spanning lighting, HVAC, pool heating, operator training and occupant-awareness campaigns. The objectives of the project were to reduce energy costs.

Factors Affecting M&V Design The owner wished to record annual cost avoidance relative to the conditions and energy usage rate of 2005 as the *baseline*. The school contained a pool and cafeteria. The school is in use year round, though it closes for a total of 5 weeks a year between sessions. The community uses the building most evenings.

The building's energy requirement is significantly affected by ambient temperature. Temperature data can be easily obtained from a nearby government weather office. No other significant energy-governing variable could be quantified.

²⁸ The annual oil and money *savings* are expressed conservatively with three *significant digits*, the lowest number of digits used in the computations as found in the efficiency tests.

Only administration offices have mechanical air-conditioning equipment, which operates for 3 months of the year.

Expected annual *savings* on the gas meter are 2,800 mcf, and 380,000 kWh on the main electricity meter.

M&V Plan An *M&V Plan* was developed showing that IPMVP Volume I, EVO 10000 – 1:2012, Option C was to be used for *savings* determination because total facility energy cost was the focus. Option C was also chosen because many *ECMs* were involved, some of which could not be directly measured.

Since *savings* are to be reported as “cost avoidance,” i.e. under *reporting period* conditions, Equation 1b) will be used.

An outline of key elements in the *M&V Plan* is shown below. Details, data and analysis are shown on the EVO subscribers’ website (www.evo-world.org).

- The *measurement boundary* of this *savings* determination was defined as:
 - An electricity account, including demand, serving the main building,
 - An auxiliary electrical account, without demand, serving lighting in the field house,
 - A natural gas account for the main building.
- The 2005 *baseline* conditions were recorded, including a strategy for the engineering department to easily capture information about future changes.
- The *baseline period*’s energy data and weather data were recorded and analyzed by simple linear regression of monthly energy use and energy demand against *degree days*. *Degree-day* data was with the base temperature, which yielded the best R^2 from a number of regression analyses performed over a range of plausible base temperatures.
- Preliminary analysis found clear correlations with weather for winter gas use and winter electricity consumption on the main meter. Analysis also showed that there is no significant weather correlation with electric demand, summer gas or electricity use. It was decided that regression would only be performed on billing periods with more than 50 *heating degree days* (HDD). It was also decided that for *reporting periods* with 50 or fewer HDDs, *adjusted-baseline* values would be derived directly from the corresponding *baseline* month, adjusted solely for the number of days in the period.

The *energy*/HDD relationships were derived for the heating season on all three accounts as shown in Table A-7-1, along with key regression statistics and coefficients where significant relationships were found.

Table A-7-1
Regression
Analysis

	Gas	Electricity		
		Main Building		Field House
		Consumption	Demand	Consumption
Units	mcf	kWh	kW	kWh
Number of months with more than 50 HDD	8	8	8	9
HDD Base	60°F	62°F	62°F	68°F
Regression Statistics:				
R^2	0.93	0.81	0.51	0.29
Standard Error of the estimate	91	15,933		
t statistic of the HDD coefficient	8.7	5.0	2.5	1.7

Assessment of Regression Analysis	Good	Good	Good	Marginal
Regression Coefficients (where accepted):				
Intercept	446.73	102,425		
HDD coefficient	1.9788	179.3916		

The regression statistics for the gas consumption and main electricity consumption show acceptable correlation with HDD as indicated by the high R^2 , and the HDD t-statistics being well above the critical IPMVP Table B-1 value of 1.89 for 8 data points and 90% confidence. These observations are logical since the primary use of gas is for building heating. There is also a significant amount of electric heat in the main building.

The field house account showed a poor t-statistic and R^2 . The building has no installed heating but must be lit longer in months of less daylight, which are also colder months. Monthly electricity use could be expected to follow a reasonably regular annual pattern related to daylight hours and occupancy, not driven by ambient temperature. Therefore the minimal correlation of this meter with HDD is ignored, and there will be no weather adjustments made to it. Instead, each month's savings report will take its *baseline energy* from the corresponding *baseline* month's consumption, adjusting for the number of days in the *reporting period*.

The main electrical-demand meter showed a poor correlation with the coldest day's weather. Therefore each month's savings report will take its *baseline* demand from the corresponding *baseline* month's actual demand, without adjustment.

- The long term impact on *savings* reports of these regression statistics was analyzed. The relative precision in winter *savings* reports will be less than $\pm 10\%$ for gas and less than $\pm 20\%$ for the main electricity account. The expected *savings* will be statistically significant for winter months since they will be more than twice the *standard error* of the *baseline* formulae (see criterion in Appendix B-1.2). The school officials felt comfortable with this expected quantified *precision*, and with possible unquantifiable errors related to simply adjusting for metering period lengths in months with 50 or fewer HDD.
- The utility rates to be used in valuing *savings* will be the then current full-price schedule appropriate for each account.

Results The *reporting-period* data for the first year was taken directly from utility bills without adjustment, and from government weather reports. This data and the calculations for the *savings* in energy and demand units, using Equation 1b), are shown on EVO's website.

Each month's current utility rate schedule was applied to each account's *adjusted-baseline energy* and *reporting-period energy* to compute *savings*. Since the gas rate changed in month 9 and the electric rate changed in month 7, two different price schedules were used for each commodity during the 12-month *savings* report. These computations are also detailed on the EVO website.

A-7.1 Whole-Facility Energy Accounting Relative To Budget

Situation The energy manager of a chain of hotels was required to annually prepare an energy budget, and routinely account for variances from budget.

Factors Affecting M&V Design Hotel guest-room occupancy, convention-area usage and weather significantly affect energy use. In order to account for energy use, the energy manager realized she needed to use M&V style techniques to adjust for these significant factors.

M&V Plan The energy manager followed IPMVP Volume I, EVO 10000 – 1:2012, Option C, since she needed to explain budget variances in management accounting reports. She always

stated her energy budgets under long-term average weather conditions and the previous year's occupancy.

Results In order to account for budget variances, as soon as a year was complete, the energy manager prepared a regression model of the usage on each utility account, using actual weather and occupancy factors for that year. She then took three steps to separately determine the primary effects of weather, occupancy and utility rates:

- Weather She inserted normal weather statistics into the most recent year's models. Using actual utility rates for the year, she determined how much the energy (and cost) would have been if the weather had been normal. (She also noted how much the actual heating and cooling *degree days* varied from normal, and from the previous year, at each location.)
- Occupancy She inserted the occupancy factors of the previous year into the most recent year's models. Using actual utility rates for the recent year, she determined how much the energy (and cost) would have been if the occupancy had been the same as the previous year. (She also noted how much the occupancy had changed from year to year at each location.)
- Utility Rates She applied the previous year's utility rate to the most recent year's consumption (and demand) to determine how much of the budget variance was related to rate changes for each utility at each location.

With the impact of these three known variables defined, the energy manager still needed to account for the remaining variances. So she inserted the recent year's weather and occupancy factors into the mathematical models of the previous year, and using current utility rates reported cost avoidance from the previous year's pattern. This cost avoidance was then analyzed in relation to changes in *static factors* recorded for each site relative to the previous year's record. All remaining variance was reported as truly random, or unknown phenomena.

This analysis process not only allowed the energy manager to account for budget variances, it also informed her of where to focus efforts to manage unaccounted variances. In addition it allowed her to make more informed budgets for subsequent years.

A-8 Multiple ECMs In A Building Without Energy Meters In The *Baseline Period* – Option D

Situation An energy efficiency project was implemented in an American university library building, involving seven ECMs spanning lighting, HVAC, operator training and occupant awareness campaigns. The building is part of a multiple-building campus without individual building meters. The objectives of the project were to reduce energy costs in the library.

Factors Affecting The M&V Design Since the project at the library was very small relative to the entire campus, its effect could not be measured using the main campus utility meters.

The university wished to achieve *savings* as quickly as possible, despite the lack of a *baseline* energy record.

Savings are to be reported continuously, as soon as possible after retrofit, using the then current energy contract prices.

M&V Plan It was decided not to wait to obtain a year's worth of energy data from new meters before implementing the measures. Instead IPMVP Volume I, EVO 10000 – 1:2012, Option D, Equation 1f) would be used, simulating pre-retrofit performance. Therefore, as part of the energy-management program steam, electricity and electric demand meters were installed on the main supply lines to the library.

The *measurement boundary* of this project was defined as all energy-using systems in the library. However the important *energy* effect was at the main campus utility meters. To

transform *energy* measured at the library to its actual impact on the campus utility bills, the following assumptions were made:

- A pound of steam at the library requires 1.5 ft³ of natural gas at the campus heating plant's gas meter. There is a fixed component in the gas use of the central plant, arising from the standing losses of the steam system. The 1.5 ft³ factor, an annual average of gas use per pound of steam produced, allocates a load-based share of this fixed component to the library.
- Electricity use at the library requires 3% more electricity at the campus electricity meter because of estimated campus transformer and distribution losses.
- Peak electric demand at the library is assumed to be coincident with the time of peak demand at the campus meter.

The expected *savings* of the *ECMs* were predicted by computer simulation with the publicly available DOE 2.1 software. A full survey of the building's systems and occupancy was needed to gather all the input data. The power requirements of five variable-air-volume air-handling systems were logged for one week to define some of the input data for this planning simulation. The simulation used long-term normal weather conditions and the occupancy and other building characteristics that prevailed at the time of the prediction. It was decided to report actual *savings* under the same conditions.

The university's gas supply contract has a marginal unit price of US\$6.25/mcf. It also has a minimum consumption level, which is only 5,300 mcf below the actual gas usage during the *baseline period*. If consumption drops by more than 5,300 mcf, the university will pay for the contract minimum amount. The contract will be renegotiated based on the results determined from this library project. The marginal electricity price at the campus meter is \$0.18/kWh in peak periods, \$0.05/kWh in off peak periods and demand is priced at \$10.25/kW-month.

Following the first year, the first year's meter data will be used as a *baseline* for a new Option C approach for this building.

Results The following steps were used to compute *savings*.

1. The new meters were calibrated and installed. Operating staff recorded monthly energy and demand for 12 months throughout the first year after *ECM* commissioning.
2. Then the original planning *simulation model* was refined to match: the *ECMs* as installed, the weather, the occupancy, and the operating profiles of the *reporting period*. The resultant simulation of space temperatures and humidities were examined to ensure they reasonably matched the typical range of indoor conditions during occupied and unoccupied days. Initially the simulation result did not match actual energy use very well, so the M&V team investigated the site further. During these additional investigations the team found that unoccupied night periods experienced very little indoor temperature change. Therefore they changed the thermal-mass characteristics of the computer model. After this correction the modeled monthly results were compared to the monthly calibration data. The highest *CV(RMSE)* of the differences was 12%, on the electric demand meter. The university felt that because these *CV(RMSE)* values met ASHRAE (2002) specifications, it could have reasonable *confidence* in the relative results of two runs of the model. Therefore this "calibrated as-built model" was archived, with both printed and electronic copy of input data, diagnostic reports and output data.
3. The calibrated as-built model was then rerun with a weather-data file corresponding to the normal year. Occupancy statistics and *static factors* were also reset to what had been observed during the *baseline period*. The resultant "**post-retrofit normal-conditions model**" was archived, with both printed and electronic copy of input data, diagnostic reports and output data.

4. The post-retrofit normal-conditions model was then adjusted to remove the *ECMs*. This “**baseline normal-conditions model**” was archived, with both printed and electronic copy of input data, diagnostic reports and output data.
5. The energy consumption of the two normal models were then compared using Equation 1f) to yield energy *savings* as shown in Table A-8-1.

**A-8-1 Simulated
Library Savings
under Normal
Conditions**

	<i>Baseline Normal Conditions Model</i>	Post-Retrofit Normal Conditions Model	<i>Savings</i>
Peak period electricity consumption (kWh)	1,003,000	656,000	347,000
Off-Peak period electricity consumption (kWh)	2,250,000	1,610,000	640,000
Electric Demand (kW-months)	7,241	6,224	1,017
Steam (thousand pounds)	12,222	5,942	6,280

6. The value of the *savings* at the campus meter were computed as shown in Table A-8-2, allowing for transformation and line losses, and contract minimum gas quantities.

**A-8-2
Campus
Savings**

	Library Energy <i>Savings</i>	Campus Energy <i>Savings</i>	Billed Energy <i>Savings</i>	Cost <i>Savings</i> US\$
Peak period electricity consumption (kWh)	347,000	357,400	357,400	64,330
Off Peak period electricity consumption (kWh)	640,000	659,200	659,200	33,000
Electric Demand (kW-months)	1,017	1,048	1,048	10,740
Steam or gas	6,280,000 pounds steam	9,420 mcf gas	5,300 mcf gas	33,000
Total				\$141,000 ²⁹

The total *savings* are shown for the year before revision to the gas contract minimum.

A-9 New Building Designed Better Than Code – Option D

Situation A new building was designed to use less energy than required by the local building code. In order to qualify for a government incentive payment, the owner was required to show that the building’s energy use during the first year of operation after commissioning and full occupancy was less than 60% of what it would have been if it had been built to code.

²⁹ The final savings number is expressed using three significant digits due to the significant figure rules in section 8.12..

Factors Affecting M&V Design Computer simulation was used extensively throughout the building design process to help meet a target energy use equal to 50% of code.

The building was built as the new corporate headquarters for a large firm. It was expected that the building would become fully occupied immediately after opening.

The owner wished to use the same energy-savings calculations that he presents to the government to show how much money was being saved as a result of his extra investment in an efficient building. He also wished to annually review variances from his initially achieved energy performance.

M&V Plan IMPVP Volume I, EVO 10000 – 1:2012, Option D will be used to demonstrate the new building's savings compared to an identical building built to building-code standards. It is possible to use either Equation 1f) comparing two simulations, or Equation 1g) comparing the simulated *baseline energy* and measured actual *energy* after correcting for calibration error. The incentive program did not specify which method should be used. The person performing the modeling felt that Equation 1f) would be more accurate. However the owner wished to use actual utility data in his final savings statement, so he required the use of Equation 1g).³⁰

Following the first year of full operation ("year one"), year one's energy and operational data will become the *baseline* for an IPMVP Volume I, EVO 10000 – 1:2012, Option C approach to reporting ongoing performance.

Results A year after commissioning and full occupancy, the original design simulation's input data was updated to reflect the as-built equipment and the current occupancy. A weather-data file was chosen from available weather files for the building's location based on the file's similarity of total heating and cooling *degree days* with year one's measured *degree days*. This similar file was appropriately adjusted to year one's actual monthly heating and cooling *degree days*. The revised input data was used to rerun the simulation.

The utility consumption data from year one was compared to this *simulation model*. After some further revisions to the simulation's input data, it was deemed that the simulation reasonably modeled the current building. This calibrated simulation was called the "as-built model."

The calibration error in the as-built model relative to actual utility data is shown in Table A-9-1.

	Gas	Electric Consumption (kWh)		Electric Demand (kW)
		Peak	Off Peak	
January	+1%	- 2%	+1%	+6%
February	- 3%	+1%	0%	- 2%
March	0%	- 2%	- 1%	- 5%
April	+2%	+3%	+1%	- 3%
May	- 2%	+5%	+2%	+6%
June	+7%	- 6%	- 2%	- 9%
July	- 6%	+2%	0%	+8%
August	+1%	- 8%	- 1%	+5%
September	- 3%	+7%	+1%	- 6%
October	- 1%	- 2%	- 1%	+5%
November	+3%	- 2%	- 1%	- 9%
December	+1%	+4%	+1%	+4%

Table A-9-1 Monthly Calibration Errors

³⁰ This method is the same as IPMVP Volume III Part 1 New Construction (2006), Option D, Method 2.

The input data for the as-built model was then changed to describe a building with the same occupancy and location but which simply meets the building-code standard. This was called the “standard model.”

The standard model’s monthly predicted energy use was adjusted by the monthly calibration errors in Table A-9-1 to yield the “**corrected-standard model**.” Actual metered data for year one was then subtracted from the corrected-standard model to yield the monthly *savings*. Percentage *savings* were computed to prove eligibility for the government incentive.

Monetary *savings* were determined for the owner by applying the then current full utility rate structure to the corrected standard model’s predicted monthly amounts. This total value was compared to the total utility payments for year one.

The year one energy data became the basis for an Option C approach for subsequent years.

APPENDIX B UNCERTAINTY

B-1 Introduction

The objective of *M&V* is to reliably determine *energy savings*. In order for *savings* reports to be reliable, they need to have a reasonable level of uncertainty. The uncertainty of a *savings* report can be managed by controlling random errors and data bias. Random errors are affected by the quality of the measurement equipment, the measurement techniques, and the design of the sampling procedure. Data bias is affected by the quality of measurement data, assumptions and analysis. Reducing errors usually increases *M&V* cost so the need for improved uncertainty should be justified by the value of the improved information (see Chapter 8.5).

Energy savings computations involve a comparison of measured *energy* data, and a calculation of “adjustments” to convert *both measurements to the* same set of operating conditions (see Chapter 4.1, Equation 1). Both the measurements and the adjustments introduce error. Errors may arise for example due to meter inaccuracy, sampling procedures or adjustment procedures. These processes produce statistical “estimates” with reported or expected values, and some level of variation. In other words, true values are not known, only estimates with some level of uncertainty. All physical measurement and statistical analysis is based on estimation of central tendencies, such as *mean* values, and quantification of variations such as range, *standard deviation*, *standard error*, and *variance*.

Statistics is the body of mathematical methods that can be applied to data to help make decisions in the face of uncertainty. For example, statistics provide ways of checking results to see if the reported *savings* are “significant,” i.e. likely to be a real effect of the ECM rather than random behavior.

Errors occur in three ways: modeling, sampling, and measurement:

- **Modeling.** Errors in mathematical modeling due to inappropriate functional form, inclusion of irrelevant variables, exclusion of relevant variables, etc. See Appendix B-2.
- **Sampling.** Sampling error arises when only a portion of the population of actual values is measured, or a biased sampling approach is used. Representation of only a portion of the population may occur in either a physical sense (i.e., only 20 of 1,000 light fixtures are metered), or in the time sense (metering occurring for only ten minutes out of every hour). See Appendix B-3.
- **Measurement.** Measurement errors arise from the accuracy of sensors, data tracking errors, drift since calibration, imprecise measurements, etc. The magnitude of such errors is largely given by manufacturer's specifications and managed by periodic re-calibration. See Appendix B-4, and Chapters 4.8.3 and 8.11.

This Appendix gives guidance on quantifying the uncertainties created by these three forms of error. Appendix B-5 discusses methods of combining quantified elements of uncertainty.

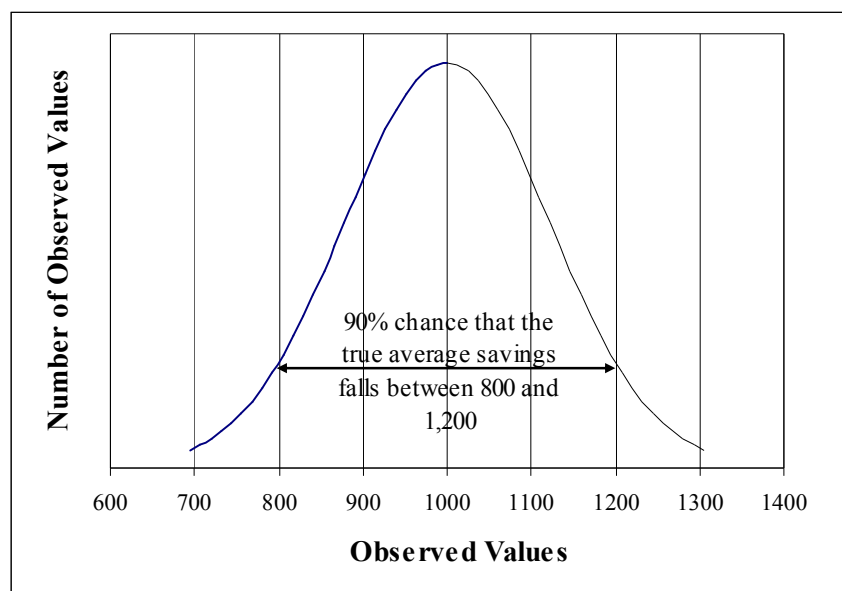
Some sources of error are unknown and unquantifiable. Examples of such sources are poor meter selection or placement, inaccurate *estimates* in Option A, or mis-estimation of *interactive effects* in Options A or B. Unknown or unquantifiable uncertainties can only be managed by following industry best practices.

An example of the use of uncertainty analysis is given in Appendix B-6. Also some of the examples in Appendix A present uncertainty calculations: A-3, A-3-2, A-4 and A-7. EVO's subscriber website (www.evo-world.org) contains details of the uncertainty calculations in A-4 and A-7.

B-1.1 Expressing Uncertainty

In order to communicate *savings* in a statistically valid manner, *savings* need to be expressed along with their associated *confidence* and *precision* levels. *Confidence* refers to the likelihood or probability that the estimated *savings* will fall within the *precision* range.³¹ For example, the *savings* estimation process may lead to a statement such as: “the best estimate of *savings* is 1,000 kWh annually (point estimate) with a 90% probability (*confidence*) that the true-average savings value falls within $\pm 20\%$ of 1,000.” A graphical presentation of this relationship is shown in Figure B-1.

Figure B-1 Normally Distributed Population



A statistical *precision* statement (the $\pm 20\%$ portion) without a *confidence* level (the 90% portion) is meaningless. The *M&V* process may yield extremely high *precision* with low *confidence*. For example, the *savings* may be stated with a *precision* of $\pm 1\%$, but the associated *confidence* level may drop from 95% to 35%.

B-1.2 Acceptable Uncertainty

Savings are deemed to be statistically valid if they are large relative to the statistical variations. Specifically, the *savings* need to be larger than twice the *standard error* (see definition in Appendix B-1.3) of the *baseline* value. If the *variance* of the *baseline* data is excessive, the unexplained random behavior in *energy* use of the *facility* or system is high, and any single *savings* determination is unreliable.

Where you cannot meet this criterion, consider using:

- more precise measurement equipment,
- more *independent variables* in any mathematical model,
- larger sample sizes, or
- an IPMVP Option that is less affected by unknown variables.

³¹ Italicized statistical terms in this Appendix are defined in Appendix B-1.3

B-1.3 Definitions of Statistical Terms

Sample Mean ($\bar{Y}$): The most widely used measure of the central tendency of a series of observations. *Sample mean* is determined by adding up the individual data points (Y_i) and dividing by the total number of these data points (n), as follows:

$$\bar{Y} = \frac{\sum Y_i}{n} \quad \text{B-1}$$

Sample Variance (S^2): *Sample variance* measures the extent to which observed values differ from each other, i.e., variability or dispersion. The greater the variability, the greater the uncertainty in the *mean*. *Sample variance*, the most important measure of variability, is found by averaging the squares of the individual deviations from the *mean*. The reason these deviations from the *mean* are squared is simply to eliminate the negative values (when a value is below the *mean*) so they do not cancel out the positive values (when a value is above the *mean*). *Sample variance* is computed as follows:

$$S^2 = \frac{\sum (Y_i - \bar{Y})^2}{n - 1} \quad \text{B-2}$$

Sample Standard Deviation (s): This is simply the square root of the *sample variance*. This brings the variability measure back to the units of the data (e.g., if the *variance* units are (kWh)², the *standard deviation* units would be kWh).

$$s = \sqrt{S^2} \quad \text{B-3}$$

Sample Standard Error (SE): This is the *sample standard deviation* divided by $\sqrt{n}$. This measure is used in estimating *precision of a sample mean*. It is also denoted as $\bar{s}$, or the "sample standard deviation of the mean" in most statistics textbooks.

$$SE = \frac{s}{\sqrt{n}} \quad \text{B-4}$$

Sample Standard Deviation of the Total (s_{tot}): Many times we are interested in the statistical properties of a *total* rather than a *mean*. The *sample standard deviation of a total* is used to define the precision about a sample total. It is defined as the square root of the sample size, $\sqrt{n}$, times the sample standard deviation:

$$s_{\text{tot}} = \sqrt{n} \cdot s \quad \text{B-5}$$

Coefficient of Variation (cv): The coefficient of variation is simply the standard deviation of a distribution expressed as a percentage of the mean. For instance, the cv of a sample total

would be the $[s_{\text{tot}}] \div [\text{sample total}]$; the cv of a sample mean would be the $[SE_{\bar{y}}] \div [\text{sample mean}]$; etc. The general formula is:

$$cv = \frac{s}{\bar{Y}}$$

B-6

Precision: *Precision* is the measure of the *absolute* or *relative* range within which the true value is expected to occur with some specified level of *confidence*. *Confidence* level refers to the probability that the quoted range contains the estimated parameter.

Absolute precision is computed from *sample standard error* using a “*t*” value from the “*t*-distribution” Table B-1:

$$t \times SE_{\bar{y}}$$

B-7

Table B.1
t-table

Degrees of Freedom DF	Confidence Level				Degrees of Freedom DF	Confidence Level			
	95%	90%	80%	50%		95%	90%	80%	50%
1	12.71	6.31	3.08	1.00	16	2.12	1.75	1.34	0.69
2	4.30	2.92	1.89	0.82	17	2.11	1.74	1.33	0.69
3	3.18	2.35	1.64	0.76	18	2.10	1.73	1.33	0.69
4	2.78	2.13	1.53	0.74	19	2.09	1.73	1.33	0.69
5	2.57	2.02	1.48	0.73	21	2.08	1.72	1.32	0.69
6	2.45	1.94	1.44	0.72	23	2.07	1.71	1.32	0.69
7	2.36	1.89	1.41	0.71	25	2.06	1.71	1.32	0.68
8	2.31	1.86	1.40	0.71	27	2.05	1.70	1.31	0.68
9	2.26	1.83	1.38	0.70	31	2.04	1.70	1.31	0.68
10	2.23	1.81	1.37	0.70	35	2.03	1.69	1.31	0.68
11	2.20	1.80	1.36	0.70	41	2.02	1.68	1.30	0.68
12	2.18	1.78	1.36	0.70	49	2.01	1.68	1.30	0.68
13	2.16	1.77	1.35	0.69	60	2.00	1.67	1.30	0.68
14	2.14	1.76	1.35	0.69	120	1.98	1.66	1.29	0.68
15	2.13	1.75	1.34	0.69	∞	1.96	1.64	1.28	0.67

Note: Calculate DF using the following,
 • DF = n - 1 (for a sample distribution)
 • DF = n - p - 1 (for a regression model)

Where,

n = sample size

p = # regression model variables

In general the true value of any statistical estimate is expected, with a given *confidence* level, to fall with the range defined by

$$\text{Range} = \text{estimate} \pm \text{absolute precision}$$

B-8

Where “estimate” is any empirically derived value of a parameter of interest (e.g., total consumption, average number of units produced).

Relative precision is the *absolute precision* divided by the estimate:

$$\frac{t * SE}{\text{Estimate}}$$

B-9

See an example use of *relative precision* in Appendix A-3. As an example of the use of these terms, consider the data in Table B-2 from 12 monthly readings of a meter, and related analysis of the difference between each reading and the *mean* of the readings (1,000):

	Actual	Computed Differences From the Mean	
	Reading	Raw	Squared
1	950	-50	2,500
2	1,090	90	8,100
3	850	-150	22,500
4	920	-80	6,400
5	1,120	120	14,400
6	820	-180	32,400
7	760	-240	57,600
8	1,210	210	44,100
9	1,040	40	1,600
10	930	-70	4,900
11	1,110	110	12,100
12	1,200	200	40,000
Total	12,000		246,600

Table B-2 Example data and analysis

The *Mean* value is: $\bar{Y} = \frac{\sum Y_i}{n} = \frac{12,000}{12} = 1,000$

The *Variance* (S^2) is: $S^2 = \frac{\sum (Y_i - \bar{Y})^2}{n-1} = \frac{246,600}{12-1} = 22,418$

The *Standard Deviation* (s) is: $s = \sqrt{S^2} = \sqrt{22,418} = 150$

The *Standard Error* is: $SE = \frac{s}{\sqrt{n}} = \frac{150}{\sqrt{12}} = 43$

Table B-1 shows that “t” is 1.80 for 12 data points (DF = 11) and a 90% confidence level. Therefore:

the *Absolute Precision* is: $t \times SE = 1.80 \times 43 = 77$ and

the *Relative Precision* is: $\frac{t \times SE}{estimate} = \frac{77}{1,000} = 7.7\%$

So, there is 90% *confidence* that the true *mean*-monthly consumption lies in the range between 923 and 1,077 kWh. It can be said with 90% *confidence* that the *mean* value of the 12 observations is $1,000 \pm 7.7\%$. Similarly it could be said:

- with 95% *confidence* that the *mean* value of the 12 observations is $1,000 \pm 9.5\%$, or
- with 80% *confidence* that the *mean* value of the 12 observations is $1,000 \pm 5.8\%$, or
- with 50% *confidence* that the *mean* value of the 12 observations is $1,000 \pm 3.0\%$.

B-2 Modeling

Mathematical modeling is used in *M&V* to prepare the *routine-adjustments* term in the various versions of Equation 1 in Chapter 4. Modeling involves finding a mathematical relationship between dependent and *independent variables*. The dependent variable, usually *energy*, is modeled as being governed by one or more *independent variable(s)* X_i , (also known as ‘explanatory’ variables). This type of modeling is called *regression analysis*.

In *regression analysis*, the model attempts to “explain” the variation in *energy* resulting from variations in the individual *independent variables* (X_i). For example, if one of the X ’s is production level, the model would assess whether the variation of *energy* from its *mean* is caused by changes in production level. The model quantifies the causation. For example, when production increases by one unit, energy consumption increases by “ b ” units, where “ b ” is called the regression coefficient.

The most common models are linear regressions of the form:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p + e$$

where:

- Y is the dependent variable, usually in the form of *energy* use during a specific time period (e.g., 30 days, 1 week, 1 day, 1 hour, etc.)
- X_{it} ($i = 1, 2, 3, \dots p$) represents the ‘ p ’ *independent variables* such as weather, production, occupancy, metering period length, etc.
- b_i ($i = 0, 1, 2, \dots p$) represents the coefficients derived for each *independent variable*, and one fixed coefficient (b_0) unrelated to the *independent variables*
- e represents the residual errors that remain unexplained after accounting for the impact of the various *independent variables*. *Regression analysis* finds the set of b_i values that minimizes the sum of squared residual-error terms (thus regression models are also called least-squares models).

An example of the above model for a building’s *energy* use is:

$$\text{monthly energy consumption} = 342,000 + (63 \times \text{HDD}) + (103 \times \text{CDD}) + (222 \times \text{Occupancy})$$

HDD and CDD are heating and cooling *degree days*, respectively. Occupancy is a measure of percent occupancy in the building. In this model 342,000 is an estimate of baseload in kWh, 63 measures the change in consumption for one additional HDD, 103 measures the change in consumption for one additional CDD, and 222 measures the change in consumption per 1% change in occupancy.

Appendix B-6 presents an example of a *regression analysis* report for a single *independent variable*, from common spreadsheet software.

B-2.1 Modeling Errors

When using regression models, as described above, several types of errors may be introduced as listed below.

1. The model is built on values that are outside the probable range of the variables to be used. A mathematical model should only be constructed using reasonable values of the dependent and *independent variables*.
2. The mathematical model may not include relevant *independent variables*, introducing the possibility of biased relationships (omitted variable bias).
3. The model may include some variables that are irrelevant.
4. The model may use inappropriate functional form.
5. The model may be based on insufficient or unrepresentative data.

Each of these types of modeling errors is discussed below.

B-2.1.1 Using Out of Range Data

If the model is built on data that are not representative of the normal *energy* behavior of the facility, then the predictions may not be relied upon. This may include inclusion of outliers, or values that are well outside the range of reasonableness. Data should be screened before building the model.

B-2.1.2 Omission of Relevant Variables

In *M&V*, *regression analysis* is used to account for changes in *energy* use. Most complex *energy* using systems are affected by innumerable *independent variables*. Regression models cannot hope to include all *independent variables*. Even if it were possible, the model would be too complex to be useful and would require excessive data gathering activities. The practical approach is to include only *independent variable(s)* thought to significantly impact *energy*.

Omission of a relevant *independent variable* may be an important error. The example model in Appendix B-2 tries to explain the variations in monthly *energy* use using several *X* variables. If a relevant *independent variable* is missing (e.g., HDD), then the model will fail to account for a significant portion of the variation in *energy*. The deficient model will also attribute some of the variation that is due to the missing variable to the variable(s) that are included in the model. The effect will be a less accurate model. .

There are no obvious indications of this problem in the standard statistical tests (except maybe a low R^2 , see B-2.2.1 below). Experience and knowledge of the engineering of the system whose performance is being measured is most valuable here.

There may be cases where a relationship is known to exist with a variable recorded during the *baseline period*. However the variable is not included in the model due to lack of budget to continue to gather the data in the *reporting period*. Such omission of a relevant variable should be noted and justified in the *M&V Plan*.

B-2.1.3 Inclusion of Irrelevant Variables

Sometimes models include irrelevant *independent variable(s)*. If the irrelevant variable has no relationship (correlation) with the included relevant variables, then it will have minimal impact on the model. However, if the irrelevant variable is correlated with other relevant variables in the model, it may bias the coefficients of the relevant variables.

Use caution in adding more *independent variables* into a *regression analysis* just because they are available. To judge the relevance of independent variables requires both experience and intuition. However, the associated *t-statistic* (see B-2.2.3 below) is one way of confirming the relevance of particular *independent variables* included in a model. Experience in *energy* analysis for the type of facility involved in any *M&V* program is necessary to determine the relevance of independent variables.

B-2.1.4 Functional Form

It is possible to model a relationship using the incorrect functional form. For example, a linear relationship might be incorrectly used in modeling an underlying physical relationship that is non-linear. For example, electricity consumption and ambient temperature tend to have a non-linear (often 'U' shaped) relationship with outdoor temperature over a one-year period in buildings that are both heated and cooled electrically. (Electricity use is high for both low and high ambient temperatures, while relatively low in mid seasons.) Modeling this non-linear relationship with a single linear model would introduce unnecessary error. Instead, separate linear models should be derived for each season.

It may also be appropriate to try higher order relationships, e.g., $Y = f(X, X^2, X^3)$.

The modeler needs to assess different functional forms and select the most appropriate among them using the evaluation measures presented in Appendix B-2.2, below.

B-2.1.5 Data Shortage

Errors may also occur from insufficient data either in terms of quantity (i.e., too few data points) or time (e.g., using summer months in the model and trying to extrapolate to winter months). The data used in modeling should be representative of the range of operations of the facility. The time period covered by the model needs to include various possible seasons, types of use, etc. This may call for either extension of the time periods used or increasing sample sizes.

B-2.2 Evaluating Regression Models

In order to evaluate how well a particular regression model explains the relationship between *energy* use and *independent variable(s)*, three tests may be performed as described below. Appendix B-6 provides evaluation of an example regression model.

B-2.2.1 Coefficient of Determination (R^2)

The first step in assessing the accuracy of a model is to examine the Coefficient of Determination, R^2 , a measure of the extent to which variations in the dependent variable *Y* from its *mean* value are explained by the regression model. Mathematically, R^2 is:

$$R^2 = \frac{\text{Explained Variation in } Y}{\text{Total Variation in } Y}$$

or more explicitly:

$$R^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} \quad \text{B-10}$$

where:

- $\hat{Y}_i$ = model predicted *energy* value for a particular data point using the measured value of the *independent variable* (i.e., obtained by plugging the *X* values into the regression model)
- $\bar{Y}$ = *mean* of the *n* measured *energy* values, found using equation B-1
- Y_i = actual observed (e.g., using a meter) value of *energy*

All statistical packages and spreadsheet *regression-analysis* tools compute the value of R^2 .

The range of possible values for R^2 is 0.0 to 1.0. An R^2 of 0.0 means none of the variation is explained by the model, therefore the model provides no guidance in understanding the variations in *Y* (i.e., the selected *independent variable(s)* give no explanation of the causes of the observed variations in *Y*). On the other hand, an R^2 of 1.0 means the model explains 100% of the variations in *Y*, (i.e., the model predicts *Y* with total certainty, for any given set of values of the *independent variable(s)*). Neither of these limiting values of R^2 is likely with real data.

In general, the greater the coefficient of determination, the better the model describes the relationship of the *independent variables* and the dependent variable. Though there is no universal standard for a minimum acceptable R^2 value, 0.75 is often considered a reasonable indicator of a good causal relationship amongst the *energy* and *independent variables*.

The R^2 test should only be used as an initial check. Models should not be rejected or accepted solely on the basis of R^2 . Finally, a low R^2 is an indication that some relevant variable(s) are not included, or that the functional form of the model (e.g., linear) is not appropriate. In this situation it would be logical to consider additional *independent variables* or a different functional form.

B-2.2.2 Standard Error of the Estimate

When a model is used to predict an *energy* value (*Y*) for given *independent variable(s)*, the accuracy of the prediction is measured by the *standard error of the estimate* ($SE_{\hat{Y}}$). This accuracy measure is provided by all standard regression packages and spreadsheets.

Once the value(s) of *independent variable(s)* are plugged into the regression model to estimate an *energy* value ($\hat{Y}$), an approximation of the range of possible values for $\hat{Y}$ can be computed using equation B-8 as:

$$\hat{Y} \pm t \times SE_{\hat{Y}}$$

where:

- $\hat{Y}$ is the predicted value of *energy* (*Y*) from the regression model
- *t* is the value obtained from the *t*-tables (see Table B-1)
- $SE_{\hat{Y}}$ is the *standard error of the estimate* (prediction). It is computed as:

$$SE_{\hat{Y}} = \sqrt{\frac{\sum (\hat{Y}_i - Y_i)^2}{n - p - 1}} \quad \text{B-11}$$

where p is the number of *independent variables* in the regression equation.
This statistic is often referred to as the root-mean squared error (*RMSE*).

Dividing the *RMSE* by the average *energy* use produces the coefficient of variation of *RMSE*, or the *CV(RMSE)*.

$$CV(RMSE) = \frac{SE_{\hat{Y}}}{\bar{Y}} \quad \text{B-12}$$

A similar measure is the mean bias error (*MBE*) defined as:

$$MBE = \frac{\sum (\hat{Y}_i - Y_i)}{n} \quad \text{B-13}$$

The *MBE* is a good indicator of overall bias in the regression estimate. Positive *MBE* indicates that regression estimates tend to overstate the actual values. Overall positive bias does tend to cancel out negative bias. The *RMSE* does not suffer from this cancellation problem.

All three measures may be used in evaluating the calibration of simulation models in Option D.

B-2.2.3 t-statistic

Since regression-model coefficients (b_k) are statistical estimates of the true relationship between an individual X variable and Y, they are subject to variation. The accuracy of the estimate is measured by the *standard error of the coefficient* and the associated value of the *t-statistic*. A *t-statistic* is a statistical test to determine whether an estimate has statistical significance. Once a value is estimated using the test, it can be compared against *critical t-values* from a *t-table* (Table B-1).

The *standard error of each coefficient* is computed by regression software. The following equation applies for the case of one independent variable.

$$SE_b = \sqrt{\frac{\sum (Y_i - \hat{Y})^2 / (n - 2)}{\sum (X_i - \bar{X})^2}} \quad \text{B-14}$$

For cases with more than one *independent variable*, the equation provides reasonable approximation when the independent variables are truly independent (i.e., not correlated). Otherwise, the equation gets very complex and the *M&V* analyst is better off using a software package to compute the *standard errors* of the coefficients.

The range within which the true value of the coefficient, b falls is found using equation B-8 as:

$$b \pm t \times SE_b$$

The *standard error of the coefficient*, b , also leads to the calculation of the *t-statistic*. This test ultimately determines if the computed coefficient is statistically significant. The *t-statistic* is computed by all statistical software using the following equation:

$$t\text{-statistic} = \frac{b}{SE_b} \quad \text{B-15}$$

Once the *t-statistic* is estimated, it can be compared against critical t values from Table B-1. If the absolute value of the *t-statistic* exceeds the appropriate number from Table B-1, then it should be concluded that the estimate is statistically valid.

A rule of thumb states that the absolute value of a *t-statistic* result of 2 or more implies that the estimated coefficient is significant relative to its *standard error*, and therefore that a relationship does exist between Y and the particular X related to the coefficient. It can then be concluded that the estimated b is not zero. However, at a *t-statistic* of about 2, the *precision* in the value of the coefficient is about $\pm 100\%$: not much of a vote of confidence in the value of b . To obtain a better precision of say $\pm 10\%$, the *t-statistic* values must be around 20, or the *standard error* of b has to be no more than 0.1 of b itself.

To improve the *t-statistic* result:

- Select *independent variable(s)* with the strongest relationship to *energy*;
- Select *independent variable(s)* whose values span the widest possible range (if X does not vary at all in the regression model, b cannot be estimated and the *t-statistic* will be poor);
- Gather and use more data points to develop the model; or
- Select a different functional form for the model; for example, one which separately determines coefficient(s) for each season in a building that is significantly affected by seasonal weather changes.

B-3 Sampling

Sampling creates errors because not all units under study are measured. The simplest sampling situation is that of randomly selecting n units from a total population of N units. In a random sample, each unit has the same probability $\left(\frac{n}{N}\right)$ of being included in the sample.

In general, the *standard error* is inversely proportional to $\sqrt{n}$. That is, increasing the sample size by a factor " f " will reduce the *standard error* (improve the precision of the estimate) by a factor of $\sqrt{f}$.

B-3.1 Sample Size Determination

You can minimize sampling error by increasing the fraction of the population that is sampled $\left(\frac{n}{N}\right)$. Increasing the sample size increases cost, of course. Several issues are critical in optimizing sample sizes. The following steps should be followed in setting the sample size.

1. **Select a homogeneous population.** In order for sampling to be cost effective, the measured units should be expected to be the same as the entire population. If there are two

different types of units in the population, they should be grouped and sampled separately. For example, when designing a sampling program to measure the operating periods of room lighting controlled by occupancy sensors, rooms occupied more or less continuously (e.g., multiple person offices) should be separately sampled from those which are only occasionally occupied (e.g., meeting rooms).

2. **Determine the desired *precision* and *confidence* levels** for the estimate (e.g., hours of use) to be reported. *Precision* refers to the error bound around the true estimate (i.e., $\pm x\%$ range around the estimate). Higher *precision* requires larger sample. *Confidence* refers to the probability that the estimate will fall in the range of *precision* (i.e., the probability that the estimate will indeed fall in the $\pm x\%$ range defined by the *precision* statement). Higher probability also requires larger samples. For example, if you want 90% *confidence* and $\pm 10\%$ *precision*, you mean that the range defined for the estimate ($\pm 10\%$) will contain the true value for the whole group (which is not observed) with a probability of 90%. As an example, in estimating the lighting hours at a *facility*, it was decided to use sampling because it was too expensive to measure the operating hours of all lighting circuits. Metering a sample of circuits provided an estimate of the true operating hours. To meet a 90/10 uncertainty criterion (*confidence* and *precision*) the sample size is determined such that, once the operating hours are estimated by sampling, the range of sample estimate ($\pm 10\%$) has to have a 90% chance of capturing the true hours of use .

The conventional approach is to design sampling to achieve a 90% *confidence* level and $\pm 10\%$ *precision*. However, the *M&V Plan* needs to consider the limits created by the budget (see Chapter 8.5). Improving *precision* from say $\pm 20\%$ to $\pm 10\%$ will increase sample size by 4 times, while improving it to $\pm 2\%$ will increase sample size by 100 times. (This is a result of the sample error being inversely proportional to $\sqrt{n}$.) Selecting the appropriate sampling criteria requires balancing accuracy requirements with *M&V* costs.

3. **Decide on the level of disaggregation.** Establish whether the *confidence* and *precision* level criteria should be applied to the measurement of all components, or to various sub-groups of components. See Appendix B-5.2. Review the *precision* and *confidence* criteria chosen in 2.
4. **Calculate Initial Sample Size.** Based on the information above, an initial estimate of the overall sample size can be determined using the following equation:

$$n_0 = \frac{z^2 * cv^2}{e^2}$$

B-16

where:

- n_0 is the initial estimate of the required sample size, before sampling begins
- cv is the *coefficient of variance*, defined as the *standard deviation* of the readings divided by the *mean*. Until the actual *mean* and *standard deviation* of the population can be estimated from actual samples, 0.5 may be used as an initial estimate for cv .
- e is the desired level of *precision*.
- z is the standard normal distribution value from Table B-1, with an infinite number of readings, and for the desired *confidence* level. For example z is 1.96 for a 95% *confidence* level (1.64 for 90%, 1.28 for 80%, and 0.67 for 50% *confidence*).

For example, for 90% *confidence* with 10% *precision*, and a cv of 0.5, the initial estimate of required sample size (n_0) is

$$n_o = \frac{1.64^2 \times 0.5^2}{0.1^2} = 67$$

In some cases (e.g., metering of lighting hours or use), it may be desirable to initially conduct a small sample for the sole purpose of estimating a *cv* value to assist in planning the sampling program. Also values from previous *M&V* work may be used as appropriate initial estimates of *cv*.

5. **Adjust initial sample size estimate for small populations.** The necessary sample size can be reduced if the entire population being sampled is no more than 20 times the size of the sample. For the initial sample size example, above, ($n_o = 67$), if the population (N) from which it is being sampled is only 200, the population is only 3 times the size of the sample. Therefore the “Finite Population Adjustment” can be applied. This adjustment reduces the sample size (n) as follows:

$$n = \frac{n_o N}{n_o + N}$$

B-17

Applying this finite population adjustment to the above example reduces the sample size (n) required to meet the 90%/±10% criterion to 50. See an example use of this adjustment in Appendix A-3-1.

6. **Finalize Sample Size.** Because the initial sample size (n_o) is determined using an assumed *cv*, it is critical to remember that the actual *cv* of the population being sampled may be different. Therefore a different actual sample size may be needed to meet the *precision* criterion. If the actual *cv* turns out to be less than the initial assumption in step 4, the required sample size will be unnecessarily large to meet the *precision* goals. If the actual *cv* turns out to be larger than assumed, then the *precision* goal will not be met unless the sample size increases beyond the value computed by Equations B-16 and B-17.

As sampling continues, the *mean* and *standard deviation* of the readings should be computed. The actual *cv* and required sample size (Equations B-16 and B-17) should be re-computed. This re-computation may allow early curtailment of the sampling process. It may also lead to a requirement to conduct more sampling than originally planned. To maintain *M&V* costs within budget it may be appropriate to establish a maximum sample size. If this maximum is actually reached after the above re-computations, the *savings* report(s) should note the actual *precision* achieved by the sampling.

B-4 Metering

Energy quantities and *independent variables* are often measured as part of an *M&V* program, using meters. No meter is 100% accurate, though more sophisticated meters may increase the accuracy towards 100%. The accuracy of selected meters is published by the meter manufacturer, from laboratory tests. Proper meter sizing, for the range of possible quantities to be measured, ensures that collected data fall within known and acceptable error limits (or *precision*).

Manufacturers typically rate *precision* as either a fraction of the current reading or as a fraction of the maximum reading on the meter’s scale. In this latter case it is important to consider where the typical readings fall on the meter’s scale before computing the *precision* of typical readings.

Over-sizing of meters whose *precision* is stated relative to maximum reading will significantly reduce the *precision* of the actual metering.

The readings of many meter systems will ‘drift’ over time due to mechanical wear. Periodic re-calibration against a known standard is required to adjust for this drift. It is important to maintain the *precision* of meters in the field through routine maintenance, and calibration against known standards.

In addition to accuracy of the meter element itself, other possibly unknown effects can reduce meter system *precision*:

- poor placement of the meter so it does not get a representative ‘view’ of the quantity it is supposed to measure (e.g., a fluid flow meter’s readings are affected by proximity to an elbow in the pipe)
- data telemetry errors which randomly or systematically clip off meter data

As a result of such unquantifiable metering errors, it is important to realize that manufacturer-quoted *precision* probably overstates the *precision* of the actual readings in the field. However there is no way to quantify these other effects.

Manufacturer *precision* statements should be in accordance with the relevant industry standard for their product. Care should be taken to determine the *confidence* level used in quoting a meter’s *precision*. Unless stated otherwise, the *confidence* is likely to be 95%.

When a single measurement is used in a *savings* computation, rather than the *mean* of several measurements, the methods of Appendix B-5 are used to combine uncertainties of several components. The *standard error* of the measured value is:

$$SE = \frac{\text{meter relative precision} \times \text{measured value}}{t}$$

B-18

Where *t* is based on the large sampling done by the meter manufacturer when developing its relative precision statement. Therefore the Table B-1 value of *t* should be for infinite sample sizes.

When making multiple readings with a meter, the observed values contain both meter error and variations in the phenomenon being measured. The *mean* of the readings likewise contains both effects. The *standard error* of the estimated *mean* value of the measurements is found using equation B-4.

Chapters 4.8.3 and 8.11 further discuss metering and provide references to other useful readings on metering.

B-5 Combining Components of Uncertainty

Both the measurement and adjustment components in Equation 1 of Chapter 4 can introduce uncertainty in reporting *savings*. The uncertainties in the individual components can be combined to enable overall statements of *savings*’ uncertainty. This combination can be performed by expressing the uncertainty of each component in terms of its *standard error*.

The components must be independent to use the following methods for combining uncertainties. Independence means that whatever random errors affect one of the components are unrelated to the errors affecting other components.

If the reported *savings* is the sum or difference of several independently determined components (*C*) (i.e., $Savings = C_1 \pm C_2 \pm \dots \pm C_p$), then the *standard error* of the reported *savings* can be estimated by:

$$SE(Savings) = \sqrt{SE(C_1)^2 + SE(C_2)^2 + ... + SE(C_p)^2} \quad B-19$$

For example, if *savings* are computed using Equation 1b) of Chapter 4 as the difference between the *adjusted-baseline energy* and measured *reporting-period energy*, the *standard error* of the difference (*savings*) is computed as:

$$SE(Savings) = \sqrt{SE(adjusted\ baseline)^2 + SE(reporting\ period\ energy)^2}$$

The *SE (adjusted baseline)* comes from the *standard error* of the estimate derived from Equation B-11. The *SE (reporting period energy)* comes from the meter accuracy using Equation B-18.

If the reported *savings* estimate is a product of several independently determined components (C_i) (i.e., $Savings = C_1 * C_2 * ... * C_p$), then the relative *standard error* of the *savings* is given approximately by:

$$\frac{SE(Savings)}{Savings} \approx \sqrt{\left(\frac{SE(C_1)}{C_1}\right)^2 + \left(\frac{SE(C_2)}{C_2}\right)^2 + ... + \left(\frac{SE(C_p)}{C_p}\right)^2} \quad B-20$$

A good example of this situation is the determination of lighting *savings* as:

$$Savings = \Delta \text{ Watts} \times \text{Hours}$$

If the *M&V Plan* requires measurement of hours of use, then “Hours” will be a value with a *standard error*. If the *M&V Plan* also includes measurement of the change in wattage, then ΔWatts will also be a value with a *standard error*. Relative *standard error* of *savings* will be computed using the formula above as follows:

$$\frac{SE(Savings)}{Savings} = \sqrt{\left(\frac{SE(\Delta \text{Watts})}{\Delta \text{Watts}}\right)^2 + \left(\frac{SE(Hours)}{Hours}\right)^2}$$

When a number of savings results are totaled and they all have the same *Standard Error*, equation B-5 or B-19 can be used to find the estimated *Standard Error* of the total reported.

$$\begin{aligned} \text{Total } SE(Savings) &= \sqrt{SE(savings_1)^2 + SE(savings_2)^2 + ... + SE(savings_N)^2} = \\ &= \sqrt{N} \times SE(Savings) \end{aligned}$$

Where N is the number of *savings* results with the same *Standard Error* that are added together. Once the *standard error* of the *savings* is determined from the above procedures, it is possible to make appropriate concluding statements about the relative amount of uncertainty inherent in the *savings*, using the mathematics of the standard normal distribution curve, Figure B-1, or data in Table B-1. For example, one can compute three values:

1. the *absolute* or *relative precision* of the total *saving*, for a given level of *confidence* (e.g., 90%), computed using the relevant t value from Table B-1 and Equation B-7 or B-9, respectively.
2. *Probable Error (PE)*, defined as the 50% confidence range. *Probable Error* represents the most likely amount of error. That is, it is equally likely that error will be larger or smaller than the *PE*. (ASHRAE, 1997). Table B-1 shows that 50% *confidence* level is achieved at $t = 0.67$ for samples sizes larger than 120, or 0.67 *standard errors* from the *mean* value. So the range of *probable error* in reported *savings* using Equation B-8 is $\pm 0.67 \times SE(Savings)$.
3. The 90% *Confidence Limit (CL)*, defined as the range where we are 90% certain that random effects did not produce the observed difference. From Table B-1 using Equation B-8, CL is $\pm 1.64 \times SE(Savings)$ for sample sizes larger than 120.

B-5.1 Assessing Interactions of Multiple Components of Uncertainty

Equations B-19 and B-20 for combining uncertainty components can be used to estimate how errors in one component will affect the accuracy of the overall *savings* report. *M&V* resources can then be designed to cost-effectively reduce error in reported *savings*. Such design considerations would take into account the costs and the effects on *savings precision* of possible improvements in the *precision* of each component.

Software applications written for common spreadsheet tools allow for easy assessment of the net error associated with the combination of multiple components of uncertainty, using Monte Carlo techniques. Monte Carlo analysis allows the assessment of multiple “what if” scenarios revealing a range of possible outcomes, their probability of occurring, and which component has the most effect on the final output. Such analysis identifies where resources need to be allocated to control error.

A simple illustration of “what if” analysis is presented below for a lighting retrofit. A nominally 96 watt light fixture is replaced with a nominal 64-watt fixture. If the fixture operates for 10 hours every day, the annual *savings* would be computed as:

$$Annual\ Savings = \frac{(96 - 64) \times 10 \times 365}{1,000} = 117 kWh$$

The new 64-watt fixture’s wattage is consistent and easily measured with accuracy. However there is much variation among the old-fixture wattages and among the hours of use in different locations. Old-fixture wattages and hours of use are not easily measured with certainty. Therefore the *savings* will also not be known with certainty. The *M&V* design challenge is to determine the impact on reported *savings* if the measurement of either of these uncertain quantities is in error by plausible amounts.

Figure B2 shows a sensitivity analysis of the *savings* for the two parameters, old-fixture watts, and hours of use. Each is varied by up to 30% and the impact on *savings* is shown. It can be seen that *savings* are significantly more sensitive to variation in old-fixture wattage than to hours of use. A 30% wattage error produces a 90% *savings* error, while a 30% error in operating hours produces only a 30% *savings* error.

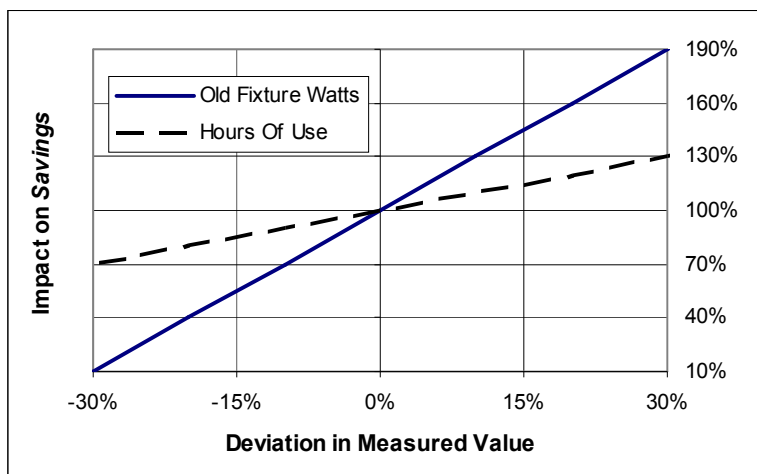


Figure B-2. Example Sensitivity Analysis – Lighting Savings

If the proposed M&V method will yield readings of old-fixture wattage with a range of uncertainty of $\pm 5\%$, the range of electricity *savings* uncertainty will be $\pm 15\%$. In other words, if the old-fixture wattage could be between 91 and 101 watts, the *savings* could be between 99 and 135 kWh annually. The range of uncertainty on the *savings* is 36 kWh (135 - 99). If the marginal value of electricity is 10 cents per kWh, the uncertainty range is about \$3.60 annually. If the wattage of the old fixture could be estimated with greater *precision* for significantly less than \$3.60, then it may be worth enhanced-measurement efforts, depending on the number of years of *savings* being considered.

Figure B2 shows that the hours-of-use term has less of an impact on final *savings* in this example (the hours-of-use line is flatter indicating lower sensitivity). It is plausible that the error in measurement of operating hours is $\pm 20\%$, so the energy-*savings* uncertainty range is also $\pm 20\%$ or ± 23 kWh (= 20% of 117 kWh). The range in *savings* is about 46 kWh (= 2 x 23 kWh), worth \$4.60 per year. Again it may be warranted to increase the accuracy in measuring the hours of use if it can be done for significantly less than \$4.60, depending upon the number of years of *savings* being considered.

The range of possible *savings* errors from errors in measuring operating hours (46 kWh) is greater than from the error in measuring the old-fixture wattages (36 kWh). This is the opposite effect from what might be expected based on the greater sensitivity of *savings* to wattage than to hours of use, as seen in Figure B2. This difference arises because the plausible error of measuring operating hours ($\pm 20\%$) is much larger than the plausible error of measuring old-fixture wattages ($\pm 5\%$).

Sensitivity analysis such as the above can take many forms. The preceding simple example was used to show the principles. Monte Carlo simulation, allows complex consideration of many different parameters, allowing M&V design to focus expenditures where most needed to improve the overall accuracy of *savings* reports.

B-5.2 Establishing Targets for Quantifiable *Savings* Uncertainty

As discussed in Appendix B-1, not all uncertainties can be quantified. However, those that can be quantified provide guidance in M&V *Planning*. By considering the M&V cost of various optional approaches to uncertainty, the M&V program can produce the type of information that is acceptable to all readers of the *savings* report, including those who have to pay for the M&V reports. Ultimately, any M&V *Plan* should report the expected level of quantifiable uncertainty (see Chapter 5).

Determination of *energy savings* requires estimating a difference in *energy* levels, rather than simply measuring the level of *energy* itself. In general, calculating a difference to suit a target *relative precision* criterion requires better *absolute precision* in the component measurements than the *absolute precision* required of the difference. For example, suppose the average load is around 500 kW, and the anticipated *savings* are around 100 kW. A $\pm 10\%$ error with 90% *confidence* (“90/10”) criterion can be applied two ways:

- If applied to the load measurements, *absolute precision* must be 50 kW (10% of 500 kW) at 90% *confidence*.
- If applied to the reported *savings*, *absolute precision* in the *savings* must be 10 kW (10% of 100 kW) at the same 90% *confidence* level. To achieve this 10 kW *absolute precision in reported savings* requires component measurement *absolute precisions* of 7 kW (using Equation B-19, if both components are to have the same *precision*).

Clearly the application of the 90/10 *confidence/precision* criterion at the level of the *savings* requires much more precision in the load measurement than a 90/10 requirement at the level of the load.

The *precision* criterion may be applied not only to *energy savings*, but also to parameters that determine *savings*. For example, suppose the *savings* amount is the product of the number (*N*) of units, hours (*H*) of operation, and change (*C*) in watts: $Savings = N \times H \times C$. The 90/10 criterion could be applied separately to each of these parameters. However, achieving 90/10 *precision* for each of these parameters separately does not imply that 90/10 is achieved for the *savings*, which is the parameter of ultimate interest. In fact using Equation B-20, the *precision* at 90% *confidence* would only be $\pm 17\%$. On the other hand, if the number of units and change in watts are assumed to be known without error, 90/10 *precision* for hours implies 90/10 *precision* for *savings*.

The *precision* standard could be imposed at various levels. The choice of level of disaggregation dramatically affects the *M&V* design and associated costs. In general, data collection requirements increase if *precision* requirements are imposed on each component. If the primary goal is to control *savings precision* for a project as a whole, it is not necessary to impose the same *precision* requirement on each component.

B-6 Example Uncertainty Analysis

To illustrate the use of the various statistical tools for uncertainty analysis, Table B-3 shows an example spreadsheet regression model output. It is a regression of a building’s 12 monthly electric-utility consumption-meter values and cooling *degree days* (CDD) over a one-year period. This is just a partial spreadsheet output. Specific values of interest are highlighted in italics.

SUMMARY OUTPUT

Table B-3
Example
Regression
Analysis
Spreadsheet
Output

Regression Statistics					
Multiple R	0.97				
R Square	0.93				
Adjusted R Square	0.92				
Standard Error	367.50				
Observations	12.00				
	Coefficients	Standard Error	T Stat	Lower 95%	Upper 95%
Intercept	5,634.15	151.96	37.08	5,295.56	5,972.74
CDD	7.94	0.68	11.64	6.42	9.45

For A Baseline of 12 Monthly kWh and CDD Data Points the derived regression model is:

$$\text{Monthly electricity consumption} = 5,634.15 + (7.94 \times \text{CDD})$$

The coefficient of determination, R^2 , (shown as “R Square” in Table B-3) is quite high at 0.93, indicating that 93% of the variation in the 12 energy data points is explained by the model using CDD data. This fact implies a very strong relationship and that the model may be used to estimate adjustment terms in the relevant form of Equation 1 in Chapter 4.

The estimated coefficient of 7.94 kWh per CDD has a *standard error* of 0.68. This SE leads to a *t-statistic* (shown as “T stat” in Table B-3) of 11.64. This *t-statistic* is then compared to the appropriate critical *t* value in Table B-1 ($t = 2.2$ for 12 data points and 95% *confidence*). Because 11.64 exceeds 2.2, CDD is a highly significant *independent variable*. The spreadsheet also shows that the range for the coefficient at the 95% level of *confidence* is 6.42 to 9.45, and implies a *relative precision* of $\pm 19\%$ ($= (7.94 - 6.42) / 7.94$). In other words, we are 95% *confident* that each additional CDD increases kWh consumption between 6.42 and 9.45 kWh.

The *standard error of the estimate* using the regression formula is 367.5. The average CDDs per month is 162 (not shown in output). To predict what electric consumption would have been under average cooling conditions, for example, this CDD value is inserted into the regression model:

$$\begin{aligned} \text{Predicted consumption} &= 5,634 + (7.94 \times 162) \\ &= 6,920 \text{ kWh per average cooling degree day month} \end{aligned}$$

Using a Table B-1 *t*-value of 2.2, for 12 data points and a 95% *confidence* level, the range of possible predictions is:

$$\text{Range of predictions} = 6,920 \pm (2.2 \times 367.5) = 6,112 \text{ to } 7,729 \text{ kWh.}$$

The *absolute precision* is approximately ± 809 kWh ($= 2.2 \times 367.5$) and the *relative precision* is $\pm 12\%$ ($= 809 / 6,920$). The spreadsheet described value for the *standard error of the estimate* provided the information needed to compute the *relative precision* expected from use of the regression model for any one month, in this case 12%.

If *reporting-period* consumption was 4,300 kWh, *savings* computed using Chapter 4, Equation 1b) will be:

$$\text{Savings} = 6,920 - 4,300 = 2,600 \text{ kWh}$$

Since the utility meter was used to obtain the *reporting-period* electricity value, its reported values may be treated as 100% accurate ($SE = 0\%$) because the utility meter defines the amounts paid, regardless of meter error. The SE of the *savings* number will be:

$$\begin{aligned} SE(\text{monthly savings}) &= \sqrt{SE(\text{adjusted baseline})^2 + SE(\text{reporting period consumption})^2} \\ &= \sqrt{367.5^2 + 0^2} = 367.5 \end{aligned}$$

Using t of 2.2, the range of possible monthly *savings* is

$$\begin{aligned} \text{Range of savings} &= 2,620 \pm (2.2 \times 367.5) \\ &= 2,620 \pm 810 = 1,810 \text{ to } 3,430 \end{aligned}$$

To determine the *precision* of the annual total of monthly *savings*, it is assumed that the *standard error* of each month's *savings* will be the same. The annual reported *savings* then have a *standard error* of:

$$SE(\text{annual savings}) = \sqrt{12 \times 367.5^2} = 1,273 \text{ kWh}$$

Since t derives from the model of the *baseline*, it remains at the 2.2 value used above. Therefore the *absolute precision* in annual *savings* is $2.2 \times 1,273 = 2,801 \text{ kWh}$.

Assuming equal monthly *savings* of 2,620 kWh, annual *savings* are 31,440 kWh, and the *relative precision* of the annual *savings* report is 9% ($= (2,801 / 31,440) \times 100$).

APPENDIX C REGION-SPECIFIC MATERIALS

This Appendix contains materials unique to the various regions of the world from which EVO has received credible contributions. These contributions may be updated separately from the rest of this Volume, so a date of publication is shown for each part. EVO encourages all regions of the world to submit materials that highlight unique aspects of M&V in their areas.

C-1 United States of America - April 2007, updated October 2011

Addition to Chapter 1.3 “IPMVP’s Relationship To Other M&V Guidelines”

- ASHRAE, Guideline 14-2002 Measurement of Energy and Demand Savings (see Reference 3 in Chapter 10). This American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc. document provides complementary detail for IPMVP. Guideline 14 had many of the same original authors as IPMVP. Though Guideline 14 provides technical detail following many of the same concepts of IPMVP, it does not use the same Option names as IPMVP. Guideline 14 is a unique and useful resource for M&V professionals around the world and is available for purchase through ASHRAE’s bookstore at <http://resourcecenter.ashrae.org/store/ashrae/>.
- M&V Guidelines: Measurement and Verification for Federal Energy Projects, Version 2.2 - 2000 (see Reference 27 in Chapter 10). The U.S. Department of Energy’s Federal Energy Management Program (FEMP) was established, in part, to reduce energy costs of operating U.S. government federal facilities. The FEMP M&V Guideline was first published in 1996 with many of the same authors as IPMVP. It provides detailed guidance on specific M&V methods for a variety of ECMs. The FEMP Guide is generally consistent with the IPMVP framework, except that it does not require site measurement of energy use for two specific ECMs. The Lawrence Berkeley National Laboratory website (<http://ateam.lbl.gov/mv/>) contains the FEMP M&V Guideline, and a number of other M&V resource documents, including one on the estimations used in Option A, and an M&V checklist.
- The U.S. State Of California’s Public Utilities Commission’s California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals (April 2006). This document provides guidance for evaluating efficiency programs implemented by a utility. It shows the role IPMVP for individual site M&V. The Protocol can be found at the California Measurement Advisory Council (CALMAC) website <http://www.calmac.org>.
- The Greenhouse Gas Protocol for Project Accounting (2005), jointly developed by the World Resources Institute and the World Business Council for Sustainable Development. The IPMVP Technical Committee was represented on the advisory committee for this document which defines means of reporting the greenhouse gas impact of carbon emission reduction and carbon sequestration projects. See www.ghgprotocol.org.

Addition to Chapters 1.2, Appendix D.6 and D.7

A widely referenced program for rating the sustainability of building designs or operations is the Leadership in Energy and Environmental Design (LEED™) of the U.S. Green Buildings Council.

Addition to Chapter 4.3, Item 6.

ORNL (1999) and ASHRAE Guideline 1-1996 define good practice in commissioning most building modifications.

Addition to Chapter 4.7, last paragraph.

ASHRAE (2002) provides more technical details on a similar retrofit isolation method

Addition to Chapter 4.7

Specific applications of retrofit-isolation techniques to common *ECMs* chosen by the United States Department of Energy are shown in Section III of FEMP (2000). Note however that FEMP's applications LE-A-01, LC-A-01 and CH-A-01 are not consistent with IPMVP because they require no measurement.

ASHRAE (2002) provides more detailed specifications for a similar method.

Addition to Chapter 4.8.1

Chapter 2.2.1 of FEMP (2000) summarizes common duties borne by parties to an *energy-performance contract*. The United States Federal Energy Management Program has also published Detailed Guidelines for FEMP M&V Option A (2002) giving further guidance on *estimation* issues faced by U.S. federal agencies. (Note: the FEMP guidelines call *estimated values* "stipulations.")

Addition to Chapter 4.9

Resources that provide detailed methods for defining new construction baselines, include: ASHRAE Standard 90.1 Appendix G – Performance Rating Method, the COMNET Commercial Building Energy Modeling Guidelines and Procedures and the Title-24 Alternative Calculation Manual. In addition, it is becoming more commonplace that commercially-available simulation programs are capable of automatically generating a minimally-compliant baseline building based on an as-designed building model per various rating methods.

ASHRAE (2002) provides more technical details on a similar method and on calibrating simulation models to utility bills.

Addition to Chapter 4.9.1

Information on building simulation programs in common use in different parts of the world can be found in Chapter 6.3 of ASHRAE (2002).

Addition to Chapter 4.9.2, Item 5.

ASHRAE (2002), Chapter 6.3, gives more information on calibration accuracy.

Addition to Chapter 4.9.1

Information on different types of building simulation models can be found in Chapter 32 of the ASHRAE Handbook (2005). The United States Department of Energy (DOE) also maintains a current list of public-domain software and proprietary building-energy-simulation programs at www.eren.doe.gov/buildings/tools_directory.

ASHRAE's simplified energy-analysis procedure may also be used if the building's heat losses, heat gains, internal loads, and HVAC systems are simple.

Other types of special-purpose programs are used to simulate *energy* use of HVAC components. See ASHRAE's HVAC02 toolkit (Brandemuehl 1993), and for boiler/chiller equipment HVAC01 toolkit (Bourdouxhe 1994a, 1994b, 1995). Simplified component air-side HVAC models are also available in a report by Knebel (1983). Equations for numerous other models have been identified as well (ASHRAE 1989, SEL 1996).

Addition to Chapter 4.9.2. Item 2

The process of obtaining and preparing actual weather data is described in depth in the *User News* Vol. 20, No. 1, which is published by Lawrence Berkeley National Laboratory and can be found at <http://gundog.lbl.gov> under Newsletters. Free actual weather data are available from U.S. D.O.E. at http://www.eere.energy.gov/buildings/energyplus/cfm/weatherdata/weather_request.cfm. Actual weather data can also be purchased. One source is the U.S. National Climatic Data Center at <http://lwf.ncdc.noaa.gov/oa/climate/climatedata.html>.

One valid method for adjusting an average weather file to resemble actual weather data is found in the WeatherMaker utility program, part of the U.S. National Renewable Energy Laboratory's software package Energy-10, available at <http://www.nrel.gov/buildings/energy10/>.

Addition to 8.3

Methods of quantifying, evaluating and reducing some of these uncertainties are discussed in this document's Appendix B, and ASHRAE (2002), section 5.2.11.³² See also Reddy & Claridge (2000) that applies standard-error-analysis methods to the typical *savings* determination.

Addition to Chapter 8.10

The U.S. Department of Energy's Building Energy Standards and Guidelines Program (BSGP), available at www.eren.doe.gov/buildings/codes_standards/buildings, provides information about U.S. residential, commercial and Federal building codes.

C-2 France - Juillet 2009

Dans tout *Plan de M&V*, l'identification de l'option choisie doit se faire au moyen de la date de publication ou du numéro de version, ainsi que de la référence du Volume de l'IPMVP, dans l'édition nationale correspondante. Exemple : IPMVP Volume I EVO 10000-1:2012:F

Chapitre 1.4

1.4A1 Benchmarks, certificats et tests régionaux

HQE : www.assohqe.org

Chapitre 4.9 Option D

4.9A1: Information relative aux différents types de modèles de simulation dans le Bâtiment

Liste des logiciels conseillés par l'ADEME (en cours d'établissement) : <http://194.117.223.129>

4.9A2 : Modèles de composants applicables

Liste des logiciels conseillés par l'ADEME (en cours d'établissement) : <http://194.117.223.129>

4.9A3 : Modèles et sources de données météorologiques applicables

Metéo-France : https://espacepro.meteofrance.com/espace_service/visite

³² It should be noted that, different from ASHRAE Guideline 14, IPMVP does not require inclusion of uncertainty reporting in savings reports.

COSTIC : <http://www.costic.com/dju/presentation.html>

4.9A4 : Méthodes de calibration applicables

Compléments méthodologiques : voir ASHRAE 2002, 1051RP

4.9A5 : Niveaux de précision minimaux recommandés

ASHRAE 2002

C-3 España - 2009

En el desarrollo del IPMVP en España, aunque no existe una normativa específica para la Medida y verificación de proyectos eficientes existen particularidades y utilidades propias de su legislación y normativa que conviene conocer.

Por ello, se anexa información específica de España:

Anexos al Capítulo 4.9.1

Para la obtención de la escala de calificación energética de edificios, en España, se ha realizado un estudio específico en el que se detalla el procedimiento utilizado para obtener los límites de dicha escala en función del tipo de edificio considerado y de la climatología de la localidad. Este procedimiento ha tomado en consideración las escalas que en la actualidad se sopesan en otros países y, en particular, la propuesta que figura en el documento del CEN prEN 15217 "Energy performance of buildings: Methods for expressing energy performance and for energy certification of buildings".

La determinación del nivel de eficiencia energética correspondiente a un edificio puede realizarse empleando dos opciones:

- La opción general, de carácter prestacional, a través de un programa informático; y la opción simplificada, de carácter prescriptivo, que desarrolla la metodología de cálculo de la calificación de eficiencia energética de una manera indirecta.
- La opción general se basa en la utilización de programas informáticos que cumplen los requisitos exigidos en la metodología de cálculo dada en el RD 47/2007. Se ha desarrollado un programa informático de referencia denominado Calener, promovido por el Ministerio de Industria, Turismo y Comercio a través del IDAE y la Dirección General de Arquitectura y Política de Vivienda del Ministerio de Vivienda.

Este programa cuenta con dos versiones:

- Calener_VYP, para edificios de Viviendas y del Pequeño y Mediano Terciario (Equipos autónomos).
- Calener_GT, para grandes edificios del sector terciario.

La utilización de programas informáticos distintos a los de referencia está sujeta a la aprobación de los mismos por parte de la Comisión Asesora para la Certificación Energética de Edificios. Esta aprobación se hará de acuerdo con los criterios que se establece en el Documento de Condiciones de Aceptación de Procedimientos Alternativos a Líder y Calener.

El Programa informático Calener es una herramienta promovida por el Ministerio de Industria, Turismo y Comercio, a través del IDAE, y por el Ministerio de Vivienda, que permite determinar el nivel de eficiencia energética correspondiente a un edificio. El programa consta de dos herramientas informáticas para una utilización más fácil por el usuario NT.

Se puede encontrar en la web del Ministerio de Industria, Turismo y Comercio, en <http://www.mityc.es/energia/desarrollo/EficienciaEnergetica/CertificacionEnergetica/ProgramaCalener/Paginas/DocumentosReconocidos.aspx>

El programa LIDER es una aplicación que permite verificar el cumplimiento de la exigencia "Limitación de la demanda energética" regulada en el DB-HE1 del nuevo Código Técnico de Edificación.

Dicho programa está incluido dentro el CALENER – VYP que se encuentra en la referencia anterior, aunque se puede obtener independientemente en la web <http://www.codigotecnico.org/index.php?id=33>

Anexos al Capítulo 4.9.2. Item 2

Los datos meteorológicos en tiempo real están disponibles en la web de la Agencia Estatal de Meteorología, dependiente del Ministerio de Medio Ambiente, Medio Rural y Marino en la web <http://www.aemet.es/es/eltiempo/observacion/ultimosdatos?k=mad>

Para la obtención de datos meteorológicos históricos igualmente en la Agencia Estatal de Meteorología, dependiente del Ministerio de Medio Ambiente, Medio Rural y Marino en la web <http://www.aemet.es/es/elclima/datosclimatologicos/resumenes>

Anexos al Capítulo 8.10

La normativa y legislación española referente al Código Técnico de la Edificación (CTE) se encuentra en la web del Ministerio de Vivienda en http://www.mviv.es/es/index.php?option=com_content&task=view&id=552&Itemid=226

Respecto al Reglamento de Instalaciones Térmicas de los Edificios (RITE) están disponibles en la web del Ministerio de Industria, Turismo y Comercio, en <http://www.mityc.es/energia/desarrollo/EficienciaEnergetica/RITE/Paginas/InstalacionesTermicas.aspx>.

C-3.1 Catalunya - Setembre 2010

Pel desenvolupament de l'IPMVP a Catalunya es podrà utilitzar tota la informació específica a la que es fa referència a l'apartat C-3 d'Espanya.

A més de la normativa europea, per fer possible la transició cap a un model energètic més sostenible s'hauran de tenir presents les normatives i decrets específics per a Catalunya, tant les que fan referència a la totalitat de les empreses com les que fan referència als edificis públics:

- [Directiva 2006/32/CE del Parlament Europeu i del Consell, de 5 d'abril de 2006 sobre l'eficiència de l'ús final de l'energia i els serveis energètics, la qual deroga la Directiva 93/76/CEE del Consell.](#)
- [Decret 21/2006, de 14 de febrer, pel qual es regula l'adopció de criteris ambientals i d'ecoeficiència en els edificis.](#) (Decret d'Ecoeficiència)

- [Directiva 2002/91/CE del Parlament Europeu i del Consell, de 16 de desembre de 2002, relativa a l'eficiència energètica dels edificis.](#)

Annex al capítol 4.9.2

Per a l'obtenció de dades meteorològiques, històriques o del any en curs, a més de la Delegació Territorial a Catalunya de l'Agència Estatal de Meteorologia (<http://www.aemet.es>) tel.: 93.221.14.72, es recomana consultar el servei meteorològic català: <http://www.meteocat.cat>

C-4 Romania - July 2010

Addition to Chapter 1.3, "IPMVP's Relationship To Other M&V Guidelines"

Another useful document for the reader of IPMVP is the Romanian National Energy Balance Elaboration Guide. The National Guide describes the way to perform an energy balance, energy audit and how to accomplish the measurement.

Addition to Chapter 4.7, "Calibration"

Devices are calibrated according to the recognized authority, the National Institute of Metrology that has as main mission the provision of scientific basis for uniformity and accuracy of measurement in Romania. Therefore calibration action must comply with its assessed laws.

Addition to Chapter 10.2, "Measurement References"

Measurements are made according to the Electric Energy Measurements Rules and Thermal Energy Measurements Rules, elaborated by ANRE (Romanian Energy Regulatory Authority) and using the Code for Electric Energy Measurement elaborated also by ANRE.

For the electric energy measurement, the rules are according to:

CEI 60044-1	Current transformers
CEI 60186	Voltage transformers
CEI 60044-2	Inductive voltage transformers
CEI 60687	Alternating current static watt-hour meters for active energy classes 0.2S and 0.5S
CEI 61036	Alternating current static watt-hour meters for active energy classes 1 and 2
CEI 61268	Alternating current static watt-hour meters for reactive energy classes 2 and 3
CEI 60521	Class 0.5, 1 and 2 alternating current watt-hour meters
CEI 60870 - 2 - 1	Telecontrol equipment and systems. Part 2: Operating conditions. Section 1: Power supply and electromagnetic compatibility.
CEI 60870 - 4	Telecontrol equipment and systems. Part 4: Performance requirements.
CEI 60870 - 5	Telecontrol equipment and systems. Part 5: Transmission protocols.
CEI 61107	Data exchange for meter reading, tariff and load control. Direct local data exchange
CEI 61334-4	Distribution automation using distribution line carrier systems. Part 4: Data communication protocols
CEI 62056-61	Electricity metering – data exchange for meter reading, tariff and load control – Part 61: Object identification system (OBIS)
CEI 62056-62	Electricity metering – data exchange for meter reading, tariff and load control – Part 62: Interface classes
CEI 62056-46	Electricity metering – data exchange for meter reading, tariff and load control – Part 46: Data Link layer using HDLC protocol

CEI 62056-53 Electricity metering – data exchange for meter reading, tariff and load control – Part 53: COSEM Application Layer
CEI 62056-21 Electricity metering – data exchange for meter reading, tariff and load control – Part 21: Direct local data exchange
CEI 62056-42 Electricity metering – data exchange for meter reading, tariff and load control – Part 42: Physical layer services and procedures for connection oriented asynchronous data exchange

For the thermal energy measurement, the rules are according to:

SR EN 1434 –1 Thermal energy meters, Part 1: General View. (1998)
STAS 6696 Taking samples (measurements) (1986)
EN 1434–2,3,4,5,6 Heat meters (1997)
ISO/IEC 7480 Information technology – Telecommunications and information exchange between systems -- Start-stop transmission signal quality at DTE/DCE interfaces (1991)
ISO/IEC 7498-1 Information technology -- Open Systems Interconnection – Basic Reference Model: The Basic Model (1994)
PE 002 Regulation for the provision and use of thermal energy (1994)
PE 003 Nomenclature of inspections, testing and proof of installation, commissioning and start-up of power plants (1984)
PE 502-8 Norms for providing technological facilities with measuring devices and automation. Heat Points (1998)
SC 001 Framework solutions for metering installation to plumbing and heating installations in existing buildings (1996)
SC 002 Framework solutions for metering water consumption, natural gas and thermal energy associated with installations from apartment blocks (1998)
OIML R 75 (International Recommendation) Thermal energy meters (1988)
NTM-3-159-94 Metrological verification of thermal energy meters (1994)

Addition to Chapter 8.7, “Data for Emission Trading”

CO₂ emissions are measured, monitored and traded according to the National Allocation Plan Regarding Greenhouse Gas Emission Certificates, that can be found at the following website: http://www.anpm.ro/Files/TEXT%20Anexe%20HG_NAP_ro-%20FINAL_20098183817246.pdf
Certificate trading is made according to EU legislation.

C-5 Bulgaria - July 2010

EU Directives – applicable in Bulgaria as references for measurement, energy efficiency, and equipment standards:

2004/22/EC	Measuring instruments
2006/95/EC	Directive 2006/95/EC of the European Parliament and of the Council of 12 December 2006 on the harmonization of the laws of Member States relating to electrical equipment designed for use within certain voltage limits (codified version)
2000/55/EC	Energy efficiency requirements for ballasts for fluorescent lighting
96/57/EC	Energy efficiency requirements for household electric

	refrigerators, freezers and combinations thereof
92/42/EEC	Efficiency requirements for new hot-water boilers fired with liquid or gaseous fuels
BDS EN 12261:2003	Gas flow meters
BDS EN 12261:2003/A1:2006	
BDS EN 12261:2003/AC:2003	
BDS EN 12405-1:2006	
BDS EN 12405-1:2006/A1:2006	
BDS EN 12480:2003	
BDS EN 12480:2003/A1:2006	
BDS EN 1359:2000	
BDS EN 1359:2000/A1:2006	
BDS EN 14154-1:2006+A1:2007	Water meters
BDS EN 14236:2009	Ultrasonic domestic gas meters
BDS EN 1434:2007	Heat meters
BDS EN 50470-1:2006	AC Electrical energy measurement
90/396/EC	Appliances burning gaseous fuels
87/404/EC	Simple pressure vessels
97/23/EC	Pressure equipment
92/75/EC	Energy labeling of household appliances
BDS EN 50294:1998/A2:2004	Lighting measurement
BDS EN 50294:2003	
BDS EN 50294:2003/A1:2003	

C-6 Czech Republic - September 2010

Referenced standards, procedures and guidelines should be replaced by European or Czech standards wherever necessary, legally required or practical. However other references in IPMVP are nevertheless informative. The most important Czech technical standards are as follows:

In the field of measuring and control tools and instruments:

- ČSN 2500 In general
- ČSN 2501 Verification of measuring instruments and measuring devices in general
- ČSN 2502 Verification of certain measuring instruments and measuring devices
- ČSN 2509 Measuring instrument accessories and record papers
- ČSN 2570 Pressure gauges in general and accessories
- ČSN 2572 Pressure gauges
- ČSN 2574 Analyzing equipments
- ČSN 2575 Volume measuring
- ČSN 2576 Volumetric weight and density measuring
- ČSN 2577 Liquid and gas flows in hollow sections measuring
- ČSN 2578 Instruments for liquid and gas flows and quantities measuring
- ČSN 2580 Thermometers in general, components
- ČSN 2581 Glass liquid thermometers
- ČSN 2582 Pressure-type thermometers, with contacts and for transformers
- ČSN 2583 Thermocouple and resistance thermometers
- ČSN 2585 Calorimeter and indicators for heating cost distribution

In the field of metrology:

ČSN 9921	Testing of ammeters, voltmeters, wattmeters
ČSN 9931	Glass thermometers
ČSN 9941	Weighing instruments
ČSN 9947	Mean absolute pressure measuring instruments
ČSN 9968	Gas flow-meters and gas volume-meters
ČSN 9971	Photometric measuring instruments
ČSN 9980	General provisions, nomenclature, symbols and units of measurement of physical-chemical properties of materials

Related to energy:

ČSN 01 1300	Legal units of measurement
ČSN 06 0210	Calculation of heat losses in buildings with central space heating
ČSN 07 0021	Hot water boilers
ČSN 07 0240	Hot water and low-pressure steam boilers
ČSN 07 0305	Evaluation of boiler losses
ČSN 07 0610	Heat exchangers water-water, steam-water
ČSN 10 5004	Compressors
ČSN 11 0010	Pumps
ČSN 12 0000	HVAC systems
ČSN 33 2000	Electrical regulations
ČSN 38 0526	Heat supply - principles
ČSN 38 5502	Gas fuels
ČSN 65 7991	Oil products, fuel oils
ČSN 73 0540	Thermal protection of buildings - parts 1, 2, 3, 4
ČSN 73 0550	Thermal properties of building structures and buildings – calculation methods
ČSN 73 0560	Thermal properties of building structures and buildings – industrial buildings
ČSN EN 835	Heat cost allocators for the determination of the consumption of room heating radiators - appliances without an electrical energy supply, based on the liquid evaporation principle
ČSN EN 834	Heat cost allocators for the determination of the consumption of room heating radiators. Appliances with electrical energy supply

Addition to Chapter 8.7, “Data for Emission Trading”

Verification of CO₂ under the EU Emission Trading Scheme must follow the relevant binding procedures set by the EU and national authorities (Act No 695/2004 Coll., as updated).

C-7 Croatia - September 2010

Addition to Chapter 4.7 “Calibration”

Replace the first sentence with: “Meters should be calibrated as recommended by the equipment manufacturer, in a laboratory approved by the Croatian agency for metering (Hrvatski zavod za mjeriteljstvo) and with a valid certificate.”

Addition to Chapter 9 “Definitions”

Baseline energy - at the end of definition add “Baseline energy consumption according to Croatian Law on efficient energy end-use” “Osnovna potrošnja energije prema Zakonu o učinkovitom korištenju energije u neposrednoj potrošnji”

Energy - at the end of definition add "See definition in Croatian Law on efficient energy end-use"
"vidi definiciju u Zakonu o učinkovitom korištenju energije u neposrednoj potrošnji"

C-8 Poland - September 2010

Requirements for measurements and measuring instruments:

- a. Ordinance of Minister of Economy on fundamental requirements for measuring instruments (Dz.U. 2007 nr 3 poz. 27; Law Gazette of 2007 No 3, item 27) and amendments (Dz.U. 2010 nr 163 poz. 1103; Law Gazette of 2010 No 163, item 1103).
- b. Law amending law of measures (Dz.U. 2010 nr 66 poz. 421; Law Gazette of 2010 No 66, item 421).
- c. Ordinance of Cabinet amending ordinance on legal units of measures (Dz.U. 2010 nr 9 poz. 61; Law Gazette of 2010 No 9, item 61)
- d. Ordinances of the Minister of Economy on meter requirements, calculation units, and testing during legal metrology inspection for:
 - gas meters: Dz.U. 2008 nr 18 poz. 115; Law Gazette of 2008 No 18, item 115
 - true energy AC electricity meters: Dz.U. 2008 nr 11 poz. 63; Law Gazette of 2008 No 11, item 63
 - liquid flow meters, other than water: Dz.U. 2008 nr 4 poz. 23; Law Gazette of 2008 No 4, item 23.

APPENDIX D USERS OF IPMVP

Though the application of IPMVP is unique to each project, certain types of users will have similar methods in their *M&V Plans* and implementation. This appendix outlines some of the key ways this document may be used by the following user groups:

- Energy performance contractors and their building customers
- Energy performance contractors and their industrial process customers
- Energy users doing their own retrofits and wanting to account for *savings*
- Facility managers properly accounting for energy budget variances
- New building designers
- New building designers seeking recognition for the sustainability of their designs
- Existing building managers seeking recognition for the environmental quality of their building operations
- Utility demand side management program designers and managers
- Water efficiency project developers
- Emission reduction trading program designers
- Energy user's seeking ISO 50001 certification

This Appendix uses terms explained in Volume I Chapters as noted in brackets, or as defined in Chapter 9 for italicized words.

D-1 Energy-Performance Contractors and Their Building Customers

The primary purpose of *M&V* in the context of building *energy-performance contracts* is presenting the actual monetary performance of a retrofit project. The *M&V Plan* becomes part of the *energy-performance contract's* terms, and defines the measurements and computations to determine payments or demonstrate compliance with a guaranteed level of performance.

M&V costs may be controlled by considering the responsibilities of all parties to the contract. Where some parameters can be estimated with accuracy sufficient for all parties, Option A (Chapter 4.8.1) may be most economical. For example, a contractor undertaking to improve the efficiency of a chiller plant may simply be required to demonstrate the before and after retrofit chiller efficiency, without consideration of the ongoing energy use, which is driven by cooling loads that are not the contractor's responsibility. However if the contractor agrees to reduce chiller plant energy use, comparison of before and after chiller plant energy use would be required. In this latter case Option B (Chapter 4.8.2) would be used if chiller plant energy meters are used, or Option C (Chapter 4.9) if whole *facility* (utility) meters are used to measure total building energy performance.

Where the *energy-performance contract* focuses on total facility performance, or where it is difficult to evaluate the effects of several *ECMs*, Option C would be used. Care should be taken to ensure that the *M&V Plan* (Chapter 5) lists baseline *static factors* and assigns responsibility for their monitoring throughout the *reporting period*. However for new construction, Option D would be used (Chapter 4.10 or IPMVP Volume III Part I). Where central metering exists on a multiple building campus and no individual building meters are yet in place, Option D (Chapter 4.10) can be used, so that the retrofit does not have to be held up to obtain new sub-meter *baseline* data for a year before planning the retrofit.

Measurements may be made throughout the term of the *energy-performance contract* or for a contract defined test period shortly after retrofit. The longer the *reporting period* (Chapter 4.5.2),

or the broader the *boundary of measurement* (Chapter 4.4), the more attention must be paid to the possibility of *baseline* change after retrofit. This possibility requires good advance recording of *static factors* in the *M&V Plan* and diligent monitoring of conditions after retrofit (Chapter 8.2).

The complexity of the *M&V* system's meter design (Chapters 4.8.3 and 8.11) and computations should also consider *M&V* costs, the expected magnitude of the *savings*, project economics and desired accuracy of reporting (Chapters 8.3 - 8.5, and Appendix B).

The prices used for monetizing the energy/water/demand units saved should be those established in the contract (Chapter 8.1).

Where an energy user feels it does not have the capability to review an *M&V Plan* or *savings* report, it may hire a verifier, separate from the *energy-performance contractor* (Chapter 8.6).

Appendix A contains examples IPMVP applications to buildings (Sections A-7, A-8, A-9, while A-2, A-3 and A-6 relate to technologies found in most buildings).

D-2 Energy-Performance Contractors and Their Industrial Customers

The primary purpose of *M&V* for industrial *energy-performance contracts* is usually to demonstrate the short-term performance of a retrofit project. Following such demonstration the plant management takes over responsibility for operation, and usually does not seek an ongoing relationship with an ESCO. The *M&V Plan* becomes part of the *energy performance contract* terms, and defines the measurements and computations to determine payments or demonstrate compliance with any guaranteed level of performance.

Industrial processes often involve complex relationships between energy use and a wider range of energy-governing variables than do buildings. In addition to weather, parameters such as product type, raw material variations, production rate and shift scheduling may be considered. Use caution selecting the *independent variables* to be used (Appendix B-2.1). The analysis becomes very difficult if attempting to identify savings at the main plant *energy* meters, especially if there is more than one product type being produced in the plant.

ECM-Isolation Options (Chapter 4.8) help to minimize the complications from production variables that are usually unrelated to terms of the performance contract. Retrofit Isolation narrows the *measurement boundary* to just those systems whose *energy* performance can be easily compared to production variables. The installation of isolation meters for *M&V* may also provide helpful feedback for process control.

M&V costs may be controlled by considering the responsibilities of all parties to the *energy-performance contract*. Where some parameters can be estimated with accuracy sufficient for all parties, Option A (Chapter 4.8.1) may be most economical. For example, a contractor who agrees to increase furnace *efficiency* may demonstrate the change in furnace energy use at peak load after installation of a flue gas heat recovery device. He is not responsible for ongoing energy use of the furnace, which is governed by production parameters beyond his control. However, if the contractor agrees instead to reduce furnace *energy use*, retrofitted furnace *energy use* is compared to the predicted energy requirements of the original furnace over a time period. In this latter case Option B (Chapter 4.8.2) governs the agreement if a meter measures the furnace's fuel consumption. Option C (Chapter 4.9) governs the agreement if main plant utility meters or departmental sub-meters measure total energy performance of the plant or a department within the plant.

Take care when using ECM-Isolation techniques to consider all energy flows affected by the *ECM* (Chapter 4.4), including *interactive effects*.

Energy-performance contracts in industrial plants often require measurements for a short *reporting period* after retrofit. Longer *reporting periods* (Chapter 4.5.2), or broader *boundaries of measurement* (Chapter 4.4), require more attention to possible *baseline* change after retrofit.

Good advance recording of *static factors* in the *M&V Plan* (Chapter 5) and diligent monitoring of conditions after retrofit (Chapter 8.2) will help identify *baseline* change.

Plant managers would normally employ long-term energy monitoring to continuously minimize energy waste. *Energy-performance contractors* may focus instead on short-term monitoring to demonstrate their performance (Chapter 4.5.2).

For retrofits that may be easily turned off temporarily, such as a heat recovery device, sequential short-term tests, which use the On/Off Test technique (Chapter 4.5.3), can demonstrate performance.

The complexity of the *M&V* system's meter design (Chapters 4.8.3 and 8.12) and computations should also consider *M&V* costs, the expected magnitude of the *savings*, project economics and desired accuracy of reporting (Chapters 8.3 - 8.5 and Appendix B).

The prices used for valuing the *savings* should be those established in the *energy performance contract* (Chapter 8.1).

The *energy* user may hire a verifier, separate from the *energy-performance contractor* (Chapter 8.6) when he or she lacks the ability to review an *M&V Plan* or *savings* report.

Appendix A contains examples of industrial applications of IPMVP (Sections A-4, A-5, while A-2, A-3.1 and A-6 relate to technologies found in most industrial plants).

D-3 Industrial and Building Energy Users Doing Their Own Retrofits

Energy users often install ECMs themselves. When they are confident about achieving the planned *savings*, a 'no *M&V*' approach leaves the entire budget available for retrofits. However energy users may need to justify investments, add credibility to requests for future investments, or quantify uncertain performance.

M&V design issues would be similar to those described in Chapters 1.4.1. or 1.4.2, above, except that there is no division of responsibility between *energy* user and an *energy performance contractor*. Reporting costs may be lower because of less formal reporting.

D-4 Facility Managers Accounting For Energy/Water Budget Variances

To manage *energy* costs successfully, a *facility* manager should understand the relationship between *energy* use and *facility* operating parameters. Important operating parameters include occupancy, production rate, and weather. If a facility manager neglects these *independent variables*, he or she may struggle to explain variances from predicted *energy* budgets. He/she also risks future budgeting errors. *Baseline adjustments* are also necessary to account for non-routine changes at the *facility*.

Even if no *savings* are planned, the calculation techniques of Chapter 4 can help to explain *energy*-budget variances. Therefore, *M&V Plans* (Chapter 5) are useful with or without retrofits. Whole facility, Option C methods (Chapter 4.9), may be used, based on main (utility) meters or sub-meters for major sections of the facility. If sub-meters are in place on specific pieces of equipment (Chapter 4.8) they may help allocate costs to user departments or tenants within the *facility* (using Option A or B approaches).

Components critical to the variations in the overall *energy* budget may be isolated for separate metering of either their *energy* use (Option B, Chapter 4.8.2) or of a key parameter of *energy* use (Option A, Chapter 4.8.1). Both of these cases call for long-term metering. Pay close attention to the cost of maintaining and calibrating meters and managing data received from the meters (see Chapter 4.8.3 and 8.12).

D-5 New Building Designers

New building investors often wish to compare their performance to what it would have been if they had not included some specific *energy* efficiency features in the design. The absence of real *baseline* data normally requires the use of Option D (Chapter 4.10) to develop a *baseline*. The computer simulation skills needed to properly apply Option D might normally be on the design team at the time of design. However the critical element of Option D is the calibration of the simulation against data gathered after a one-year period. Therefore it is important to ensure that the simulation skills remain available until calibration is achieved.

Following the first year of steady operation it would be normal to use the actual *energy* data of the first steady year as a new *baseline*, switching to using Option C (Chapter 4.9) to determine changes from the first year's new *baseline*.

All the challenges for new buildings are addressed in more depth in IPMVP Volume III Part I, New Construction, including different methods for special situations.

D-6 New Building Designers Seeking Recognition for Their Sustainable Designs

Building designers may seek to have their designs recognized under a sustainable design program. To qualify, the building may need to have an *M&V* system that adheres to IPMVP. IPMVP adherence is defined in Chapter 7 as preparation of an *M&V Plan* (Chapter 5) using IPMVP terminology, and then following the *M&V Plan*.

The designer would also follow the guidance above in Chapter 1.4.5 and IPMVP Volume III Part I.

D-7 Existing Building Managers Seeking Recognition for the Environmental Quality of their Building Operations

Managers of existing buildings may seek recognition for the environmental quality of their operating methods. To qualify they may need to have an *M&V* system that adheres to IPMVP. IPMVP adherence is defined in Chapter 7 as preparation of an *M&V Plan* (Chapter 5) using IPMVP terminology and then following the *M&V Plan*. IPMVP's Retrofit Isolation type *M&V* (Chapter 4.8) may also assist in obtaining recognition for the number of installed sub-meters.

Option C (Chapter 4.9) would provide the total facility performance monitoring that is appropriate to existing buildings. However if no whole building meters existed before seeking recognition, Option D (Chapter 4.10) is needed during the period of developing a *baseline* for a year after main meters are initially installed on the building.

Building managers would also follow the guidance above in Chapter 1.4.3.

D-8 Regional Efficiency-Program Designers and Managers

Designers and managers of regional or utility-company demand-side-management (DSM) programs usually need to develop rigorous ways to evaluate the effectiveness of their energy-efficiency programs. One way to evaluate a DSM program's impact is to assess the *savings* made in randomly chosen end-user *facilities*. This data can be used to project the results across the entire group of DSM program participants. Use IPMVP Options presented in Chapter 4 to assess the *savings* in the sample *facilities*.

The evaluation design for any regional program should specify which of the IPMVP Options are allowable. It should also specify the minimum required sampling, measurement, and analytical accuracies, in order to provide sufficient rigor in program reporting.

Utilities already have the whole facility data for their commodity in their databases, so may apply Option C (Chapter 4.9) on all program participants or a sample of them. However without adequate knowledge of changes within each facility, a large percentage of variation in the *savings* should be expected, especially as time elapses between the *baseline* and *reporting periods*.

EVO is monitoring the needs of the utility program evaluation community. EVO is considering development of special M&V guidance for DSM program evaluation, and for establishing baselines for measuring the ‘demand response’ of customers receiving utility price or curtailment signals (see Preface – EVO’s Future Plans).

D-9 Water-Efficiency Project Developers

Water-efficiency *M&V* is analogous to energy-efficiency *M&V*, so uses similar *M&V* techniques. The relevant technique for any project depends upon the nature of the change being evaluated, and upon the user’s situation as discussed in Chapters 1.4.1 through 1.4.5 and 1.4.8.

Water-consuming equipment is often in the control of facility users (building occupants or production managers). Therefore it can be difficult to monitor user behavior as needed to make adjustments to total-facility water use for the application of Option C methods. Retrofit Isolation methods are often more easily applied (Chapter 4.8), using a sample of the retrofits (Appendix B-3) to demonstrate the performance of an entire group of changes.

Where outdoor water use is being evaluated, the adjustments term in IPMVP Equation 1 (Chapter 4) may be related to parameters that drive water use such as rainfall.

Liquid flow measurements devices (see Chapter 8.11, Table 5) are most commonly applied in *M&V* for water efficiency projects.

D-10 Emission Trading Programs

Energy-efficiency programs can be central to helping large energy users meet their regulated emission allocation. All of the techniques of this document help energy users manage their energy use, through proper accounting (Chapters 1.4.3 and 1.4.4).

Energy-efficiency projects may also be the basis of trades in emission-reduction commodities (credits, offsets, set asides, etc). Since such trades must hold up under public scrutiny, compliance with an industry recognized protocol adds credibility to emission reduction claims.

Trading-program designers should specify compliance with IPMVP, 2002 edition or later. They may go further and require fully measured energy-*savings* approaches (i.e. Options B or C, Chapters 4.8.2 or 4.9). This further specification reduces the uncertainty in the quantification by eliminating Options using estimated or simulated values rather than measured values.

Chapter 8.7 discusses the special *M&V* design issues for emission trades.

D-11 Energy Users Seeking ISO 50001 Certification

The management methods required under ISO 50001 focus on managing total facility utility costs. IPMVP’s Options C and D (Chapters 4.9 and 4.10) describe methods useful for this purpose, even if no savings are being sought. However IPMVP cautions on the challenges of detecting small changes in whole facility utility energy use data. IPMVP offers both general guidance (Chapter 4.9) and specific guidance (Appendix B-1.2) on how large total facility energy use changes must be to be reportable with any statistical validity. If expected savings are not large enough, it would be appropriate to also use sub-meters for different sections of the facility. Also IPMVP’s Retrofit Isolation Options A and/or B (Chapter 4.8) would also be appropriate management tools for specific energy efficiency projects. However Chapter 4.8 cautions that Retrofit Isolation results cannot be correlated with whole facility utility bills.

INDEX

A

adherence, viii, 2, 40, 70, 120
adjustments, viii, 4, 7, 8, 11, 12, 13, 14, 15, 17, 18, 19,
 baseline adjustments, 42, 57
 non-routine adjustments, 13, 17, 23, 25, 42, 55, 57, 76, 78, 79
 routine adjustments, 12, 17, 18, 42, 43, 57
ASHRAE Guideline 14, 59, 107
avoided energy use, 7, 14, 16, 55, 57

B

baseline period, 8, 11, 17, 18, 55

C

calibration, 25, 29, 30, 38, 43, 64, 66,
confidence, 44, 55, 57,
cost, 5, 6, 37, 41, 44, 45, 46,
cv, 55, 89
CV(RMSE), 55, 96
cycle, 11, 12, 55

D

degree day, 27, 48, 55

E

energy conservation measure, 1, 4, 7, 8, 56

I

independent variable, 12, 14, 15, 26, 39, 42, 55, 57
interactive effect, 11, 18, 19, 32, 36, 43, 56

L

LEED, 3, 107

M

mean bias error, 56, 96
measurement boundary, 10, 16, 18, 19, 25, 36, 57
model, 14, 15

N

normalized savings, 15, 37, 55, 57

O

Option A, 17, 18, 19, 21, 45
Option B, 17, 18, 23, 45
Option C, 16, 18, 25, 45
Option D, 118, 28, 45

P

precision, 55, 57
price, 37, 41, 42, 56,
Principles of M&V, 6, 40
probable error, 57, 102
proxy, 16, 57, 70

R

R^2 , 57, 94
regression analysis, 18, 26, 57, 92, 93
reporting period, 7, 12, 37, 55, 56, 57

S

sample, 20, 37, 38, 43,
 sampling, 20, 23, 43, 97, 98
significant digits, 44, 52, 58
standard deviation, 58, 89, 91
standard error, 58, 87, 88, 91
 standard error of the coefficient, 96, 97
 standard error of the estimate, 101
static factor, 13, 37, 42, 43, 57, 58

T

t-statistic, 58, 94, 96, 97

U

uncertainty, 21, 37, 43, 44, 45, 87

V

variance, 28, 58, 87, 88, 91
verification, 4, 9, 10, 58



EVO thanks its current Primary Subscribers:

BC Hydro

San Diego Gas & Electric Company

Southern California Edison

Gas Natural Fenosa

Schneider Electric

Services Industriels de Genève - SIG

CALIFORNIA STANDARD PRACTICE MANUAL

ECONOMIC ANALYSIS OF DEMAND-SIDE PROGRAMS AND PROJECTS

OCTOBER 2001

Table of Contents

	Page
Chapter 1	1
Basic Methodology	1
Background.....	1
Demand-Side Management Categories and Program Definitions.....	2
Basic Methods	4
Balancing the Tests	6
Limitations: Externality Values and Policy Rules.....	6
Externality Values.....	7
Policy Rules.....	7
Chapter 2	8
Participant Test	8
Definition.....	8
Benefits and Costs.....	8
How the Results Can be Expressed.....	9
Strengths of the Participant Test.....	9
Weaknesses of the Participant Test.....	10
Formulae.....	10
Chapter 3	13
The Ratepayer Impact Measure Test	13
Definition.....	13
Benefits and Costs.....	13
How the Results can be Expressed	13
Strengths of the Ratepayer Impact Measure (RIM) Test	14
Weaknesses of the Ratepayer Impact Measure (RIM) Test	15
Chapter 4	18
Total Resource Cost Test	18
Definition.....	18
How the Results Can be Expressed.....	19
Strengths of the Total Resource Cost Test	21
Weakness of the Total Resource Cost Test	21
Formulas.....	21
Chapter 5	23
Program Administrator Cost Test.....	23
Definition.....	23
Benefits and Costs.....	23
How the Results Can be Expressed.....	23
Strengths of the Program Administrator Cost Test.....	24
Weaknesses of the Program Administrator Cost Test.....	24
Formulas.....	24
Appendix A	26
Inputs to Equations and Documentation	26
Appendix B.....	28

Summary of Equations and Glossary of Symbols	28
Basic Equations.....	28
Participant Test	28
Ratepayer Impact Measure Test.....	28
Total Resource Cost Test.....	28
Program Administrator Cost Test	28
Benefits and Costs.....	29
Participant Test	29
Ratepayer Impact Measure Test.....	29
Total Resource Cost Test.....	29
Program Administrator Cost Test	30
Glossary of Symbols	30
Appendix C.....	33
Derivation of Rim Lifecycle Revenue Impact Formula.....	33
Rate Impact Measure.....	33

Chapter 1

Basic Methodology

Background

Since the 1970s, conservation and load management programs have been promoted by the California Public Utilities Commission (CPUC) and the California Energy Commission (CEC) as alternatives to power plant construction and gas supply options. Conservation and load management (C&LM) programs have been implemented in California by the major utilities through the use of ratepayer money and by the CEC pursuant to the CEC legislative mandate to establish energy efficiency standards for new buildings and appliances.

While cost-effectiveness procedures for the CEC standards are outlined in the Public Resources Code, no such official guidelines existed for utility-sponsored programs. With the publication of the *Standard Practice for Cost-Benefit Analysis of Conservation and Load Management Programs* in February 1983, this void was substantially filled. With the informal "adoption" one year later of an appendix that identified cost-effectiveness procedures for an "All Ratepayers" test, C&LM program cost effectiveness consisted of the application of a series of tests representing a variety of perspectives-participants, non-participants, all ratepayers, society, and the utility.

The Standard Practice Manual was revised again in 1987-88. The primary changes (relative to the 1983 version), were: (1) the renaming of the "Non-Participant Test" to the "Ratepayer Impact Test"; (2) renaming the All-Ratepayer Test" to the "Total Resource Cost Test;"; (3) treating the "Societal Test" as a variant of the "Total Resource Cost Test;," and, (4) an expanded explanation of "demand-side" activities that should be subjected to standard procedures of benefit-cost analysis.

Further changes to the manual captured in this (2001) version were prompted by the cumulative effects of changes in the electric and natural gas industries and a variety of changes in California statute related to these changes. As part of the major electric industry restructuring legislation of 1996 (AB1890), for example, a public goods charge was established that ensured minimum funding levels for "cost effective conservation and energy efficiency" for the 1998-2002 period, and then (in 2000) extended through the year 2011. Additional legislation in 2000 (AB1002) established a natural gas surcharge for similar purposes. Later in that year, the Energy Security and Reliability Act of 2000 (AB970) directed the California Public Utilities Commission to establish, by the Spring of 2001, a distribution charge to provide revenues for a self generation program and a directive to consider changes to cost-effectiveness methods to better account for reliability concerns.

In the Spring of 2001, a new state agency — the Consumer Power and Conservation Financing Authority — was created. This agency is expected to provide additional revenues in the form of state revenue bonds that could supplement the amount and type of public financial resources to finance energy efficiency and self generation activities.

The modifications to the Standard Practice Manual reflect these more recent developments in several ways. First, the “Utility Cost Test” is renamed the “Program Administrator Test” to include the assessment of programs managed by other agencies. Second, a definition of self generation as a type of “demand-side” activity is included. Third, the description of the various potential elements of “externalities” in the Societal version of the TRC test is expanded. Finally the limitations section outlines the scope of this manual and elaborates upon the processes traditionally instituted by implementing agencies to adopt values for these externalities and to adopt the policy rules that accompany this manual.

Demand-Side Management Categories and Program Definitions

One important aspect of establishing standardized procedures for cost-effectiveness evaluations is the development and use of consistent definitions of categories, programs, and program elements.

This manual employs the use of general program categories that distinguish between different types of demand-side management programs, conservation, load management, fuel substitution, load building and self-generation. Conservation programs reduce electricity and/or natural gas consumption during all or significant portions of the year. ‘Conservation’ in this context includes all ‘energy efficiency improvements’. An energy efficiency improvement can be defined as reduced energy use for a comparable level of service, resulting from the installation of an energy efficiency measure or the adoption of an energy efficiency practice. Level of service may be expressed in such ways as the volume of a refrigerator, temperature levels, production output of a manufacturing facility, or lighting level per square foot. Load management programs may either reduce electricity peak demand or shift demand from on peak to non-peak periods.

Fuel substitution and load building programs share the common feature of increasing annual consumption of either electricity or natural gas relative to what would have happened in the absence of the program. This effect is accomplished in significantly different ways, by inducing the choice of one fuel over another (fuel substitution), or by increasing sales of electricity, gas, or electricity and gas (load building). Self generation refers to distributed generation (DG) installed on the customer’s side of the electric utility meter, which serves some or all of the customer’s electric load, that otherwise would have been provided by the central electric grid.

In some cases, self generation products are applied in a combined heat and power manner, in which case the heat produced by the self generation product is used on site to provide some or all of the customer’s thermal needs. Self generation technologies include, but are not limited to, photovoltaics, wind turbines, fuel cells, microturbines, small gas-fired turbines, and gas-fired internal combustion engines.

Fuel substitution and load building programs were relatively new to demand-side management in California in the late 1980s, born out of the convergence of several factors

that translated into average rates that substantially exceeded marginal costs. Proposals by utilities to implement programs that increase sales had prompted the need for additional procedures for estimating program cost effectiveness. These procedures maybe applicable in a new context. AB 970 amended the Public Utilities Code and provided the motivation to develop a cost-effectiveness method that can be used on a common basis to evaluate all programs that will remove electric load from the centralized grid, including energy efficiency, load control/demand-responsiveness programs and self-generation. Hence, self-generation was also added to the list of demand side management programs for cost-effectiveness evaluation. In some cases, self-generation programs installed with incremental load are also included since the definition of self-generation is not necessarily confined to projects that reduce electric load on the grid. For example, suppose an industrial customer installs a new facility with a peak consumption of 1.5 MW, with an integrated on-site 1.0 MW gas fired DG unit. The combined impact of the new facility is load building since the new facility can draw up to 0.5 MW from the grid, even when the DG unit is running. The proper characterization of each type of demand-side management program is essential to ensure the proper treatment of inputs and the appropriate interpretation of cost-effectiveness results.

Categorizing programs is important because in many cases the same specific device can be and should be evaluated in more than one category. For example, the promotion of an electric heat pump can and should be treated as part of a conservation program if the device is installed in lieu of a less efficient electric resistance heater. If the incentive induces the installation of an electric heat pump instead of gas space heating, however, the program needs to be considered and evaluated as a fuel substitution program. Similarly, natural gas-fired self-generation, as well as self-generation units using other non-renewable fossil fuels, must be treated as fuel-substitution. In common with other types of fuel-substitution, any costs of gas transmission and distribution, and environmental externalities, must be accounted for. In addition, cost-effectiveness analyses of self-generation should account for utility interconnection costs. Similarly, a thermal energy storage device should be treated as a load management program when the predominant effect is to shift load. If the acceptance of a utility incentive by the customer to, install the energy storage device is a decisive aspect of the customer's decision to remain an electric utility customer (i.e., to reject or defer the option of installing a gas-fired cogeneration system), then the predominant effect of the thermal energy storage device has been to substitute electricity service for the natural gas service that would have occurred in the absence of the program.

In addition to Fuel Substitution and Load Building Programs, recent utility program proposals have included reference to "load retention," "sales retention," "market retention," or "customer retention" programs. In most cases, the effect of such programs is identical to either a Fuel Substitution or a Load Building program — sales of one fuel are increased relative to sales without the program. A case may be made, however, for defining a separate category of program called "load retention." One unambiguous example of a load retention program is the situation where a program keeps a customer from relocating to another utility service area. However, computationally the equations and guidelines included in this manual to accommodate Fuel Substitution and Load Building programs can also handle this special situation as well.

Basic Methods

This manual identifies the cost and benefit components and cost-effectiveness calculation procedures from four major perspectives: Participant, Ratepayer Impact Measure (RIM), Program Administrator Cost (PAC), and Total Resource Cost (TRC). A fifth perspective, the Societal, is treated as a variation on the Total Resource Cost test. The results of each perspective can be expressed in a variety of ways, but in all cases it is necessary to calculate the net present value of program impacts over the lifecycle of those impacts.

Table I summarizes the cost-effectiveness tests addressed in this manual. For each of the perspectives, the table shows the appropriate means of expressing test results. The primary unit of measurement refers to the way of expressing test results that are considered by the staffs of the two Commissions as the most useful for summarizing and comparing demand-side management (DSM) program cost-effectiveness. Secondary indicators of cost-effectiveness represent supplemental means of expressing test results that are likely to be of particular value for certain types of proceedings, reports, or programs.

This manual does not specify how the cost-effectiveness test results are to be displayed or the level at which cost-effectiveness is to be calculated (e.g., groups of programs, individual programs, and program elements for all or some programs). It is reasonable to expect different levels and types of results for different regulatory proceedings or for different phases of the process used to establish proposed program-funding levels. For example, for summary tables in general rate case proceedings at the CPUC, the most appropriate tests may be the RIM lifecycle revenue impact, Total Resource Cost, and Program Administrator Cost test results for programs or groups of programs. The analysis and review of program proposals for the same proceeding may include Participant test results and various additional indicators of cost-effectiveness from all tests for each individual program element. In the case of cost-effectiveness evaluations conducted in the context of integrated long-term resource planning activities, such detailed examination of multiple indications of costs and benefits may be impractical.

Table I
Cost-Effectiveness Tests

Participant	
Primary	Secondary
Net present value (all participants)	Discounted payback (years) Benefit-cost ratio Net present value (average participant)
Ratepayer Impact Measure	
Lifecycle revenue impact per Unit of energy (kWh or therm) or demand customer (kW)	Lifecycle revenue impact per unit Annual revenue impact (by year, per kWh, kW, therm, or customer) First-year revenue impact (per kWh, kW, therm, or customer) Benefit-cost ratio
Net present value	
Total Resource Cost	
Net present value (NPV)	Benefit-cost ratio (BCR) Levelized cost (cents or dollars per unit of energy or demand) Societal (NPV, BCR)
Program Administrator Cost	
Net present value	Benefit-cost ratio Levelized cost (cents or dollars per unit of energy or demand)

Rather than identify the precise requirements for reporting cost-effectiveness results for all types of proceedings or reports, the approach taken in this manual is to (a) specify the components of benefits and costs for each of the major tests, (b) identify the equations to be used to express the results in acceptable ways; and (c) indicate the relative value of the different units of measurement by designating primary and secondary test results for each test.

It should be noted that for some types of demand-side management programs, meaningful cost-effectiveness analyses cannot be performed using the tests in this manual. The following guidelines are offered to clarify the appropriated "match" of different types of programs and tests:

1. For generalized information programs (e.g., when customers are provided generic information on means of reducing utility bills without the benefit of on-site evaluations or customer billing data), cost-effectiveness tests are not expected because of the extreme difficulty in establishing meaningful estimates of load impacts.

2. For any program where more than one fuel is affected, the preferred unit of measurement for the RIM test is the lifecycle revenue impacts per customer, with gas and electric components reported separately for each fuel type and for combined fuels.
3. For load building programs, only the RIM tests are expected to be applied. The Total Resource Cost and Program Administrator Cost tests are intended to identify cost-effectiveness relative to other resource options. It is inappropriate to consider increased load as an alternative to other supply options.
4. Levelized costs may be appropriate as a supplementary indicator of cost per unit for electric conservation and load management programs relative to generation options and gas conservation programs relative to gas supply options, but the levelized cost test is not applicable to fuel substitution programs (since they combine gas and electric effects) or load building programs (which increase sales).

The delineation of the various means of expressing test results in **Table 1** is not meant to discourage the continued development of additional variations for expressing cost-effectiveness. Of particular interest is the development of indicators of program cost effectiveness that can be used to assess the appropriateness of program scope (i.e. level of funding) for General Rate Case proceedings. Additional tests, if constructed from the net present worth in conformance with the equations designated in this manual, could prove useful as a means of developing methodologies that will address issues such as the optimal timing and scope of demand-side management programs in the context of overall resource planning.

Balancing the Tests

The tests set forth in this manual are not intended to be used individually or in isolation. The results of tests that measure efficiency, such as the Total Resource Cost Test, the Societal Test, and the Program Administrator Cost Test, must be compared not only to each other but also to the Ratepayer Impact Measure Test. This multi-perspective approach will require program administrators and state agencies to consider tradeoffs between the various tests. Issues related to the precise weighting of each test relative to other tests and to developing formulas for the definitive balancing of perspectives are outside the scope of this manual. The manual, however, does provide a brief description of the strengths and weaknesses of each test (Chapters 2, 3, 4, and 5) to assist users in qualitatively weighing test results.

Limitations: Externality Values and Policy Rules

The list of externalities identified in Chapter 4, page 27, in the discussion on the Societal version of the Total Resource Cost test is broad, illustrative and by no means exhaustive. Traditionally, implementing agencies have independently determined the details such as the components of the externalities, the externality values and the policy rules which specify the contexts in which the externalities and the tests are used.

Externality Values

The values for the externalities have not been provided in the manual. There are separate studies and methodologies to arrive at these values. There are also separate processes instituted by implementing agencies before such values can be adopted formally.

Policy Rules

The appropriate choice of inputs and input components vary by program area and project. For instance, low income programs are evaluated using a broader set of non-energy benefits that have not been provided in detail in this manual. Implementing agencies traditionally have had the discretion to use or to not use these inputs and/or benefits on a project- or program-specific basis. The policy rules that specify the contexts in which it is appropriate to use the externalities, their components, and tests mentioned in this manual are an integral part of any cost-effectiveness evaluation. These policy rules are not a part of this manual.

To summarize, the manual provides the methodology and the cost-benefit calculations only. The implementing agencies (such as the California Public Utilities Commission and the California Energy Commission) have traditionally utilized open public processes to incorporate the diverse views of stakeholders before adopting externality values and policy rules which are an integral part of the cost-effectiveness evaluation.

Chapter 2

Participant Test

Definition

The Participants Test is the measure of the quantifiable benefits and costs to the customer due to participation in a program. Since many customers do not base their decision to participate in a program entirely on quantifiable variables, this test cannot be a complete measure of the benefits and costs of a program to a customer.

Benefits and Costs

The benefits of participation in a demand-side program include the reduction in the customer's utility bill(s), any incentive paid by the utility or other third parties, and any federal, state, or local tax credit received. The reductions to the utility bill(s) should be calculated using the actual retail rates that would have been charged for the energy service provided (electric demand or energy or gas). Savings estimates should be based on gross savings, as opposed to net energy savings¹.

In the case of fuel substitution programs, benefits to the participant also include the avoided capital and operating costs of the equipment/appliance not chosen. For load building programs, participant benefits include an increase in productivity and/or service, which is presumably equal to or greater than the productivity/ service without participating. The inclusion of these benefits is not required for this test, but if they are included then the societal test should also be performed.

The costs to a customer of program participation are all out-of-pocket expenses incurred as a result of participating in a program, plus any increases in the customer's utility bill(s). The out-of-pocket expenses include the cost of any equipment or materials purchased, including sales tax and installation; any ongoing operation and maintenance costs; any removal costs (less salvage value); and the value of the customer's time in arranging for the installation of the measure, if significant.

¹ Gross energy savings are considered to be the savings in energy and demand seen by the participant at the meter. These are the appropriate program impacts to calculate bill reductions for the Participant Test. Net savings are assumed to be the savings that are attributable to the program. That is, net savings are gross savings minus those changes in energy use and demand that would have happened even in the absence of the program. For fuel substitution and load building programs, gross-to-net considerations account for the impacts that would have occurred in the absence of the program.

How the Results can be Expressed

The results of this test can be expressed in four ways: through a net present value per average participant, a net present value for the total program, a benefit-cost ratio or discounted payback. The primary means of expressing test results is net present value for the total program; discounted payback, benefit-cost ratio, and per participant net present value are secondary tests.

The discounted payback is the number of years it takes until the cumulative discounted benefits equal or exceed the cumulative discounted costs. The shorter the discounted payback, the more attractive or beneficial the program is to the participants. Although "payback period" is often defined as undiscounted in the textbooks, a discounted payback period is used here to approximate more closely the consumer's perception of future benefits and costs.²

Net present value (NPVp) gives the net dollar benefit of the program to an average participant or to all participants discounted over some specified time period. A net present value above zero indicates that the program is beneficial to the participants under this test.

The benefit-cost ratio (BCRp) is the ratio of the total benefits of a program to the total costs discounted over some specified time period. The benefit-cost ratio gives a measure of a rough rate of return for the program to the participants and is also an indication of risk. A benefit-cost ratio above one indicates a beneficial program.

Strengths of the Participant Test

The Participants Test gives a good "first cut" of the benefit or desirability of the program to customers. This information is especially useful for voluntary programs as an indication of potential participation rates.

For programs that involve a utility incentive, the Participant Test can be used for program design considerations such as the minimum incentive level, whether incentives are really needed to induce participation, and whether changes in incentive levels will induce the desired amount of participation.

These test results can be useful for program penetration analyses and developing program participation goals, which will minimize adverse ratepayer impacts and maximize benefits.

For fuel substitution programs, the Participant Test can be used to determine whether program participation (i.e. choosing one fuel over another) will be in the long-run best interest of the customer. The primary means of establishing such assurances is the net present value, which looks at the costs and benefits of the fuel choice over the life of the equipment.

² It should be noted that if a demand-side program is beneficial to its participants ($NPVp \geq 0$ and $BCRp \geq 1.0$) using a particular discount rate, the program has an internal rate of return (IRR) of at least the value of the discount rate.

Weaknesses of the Participant Test

None of the Participant Test results (discounted payback, net present value, or benefit-cost ratio) accurately capture the complexities and diversity of customer decision-making processes for demand-side management investments. Until or unless more is known about customer attitudes and behavior, interpretations of Participant Test results continue to require considerable judgment. Participant Test results play only a supportive role in any assessment of conservation and load management programs as alternatives to supply projects.

Formulae

The following are the formulas for discounted payback, the net present value (NPVp) and the benefit-cost ratio (BCRp) for the Participant Test.

$$\begin{aligned} \text{NPV}_p &= B_p - C_p \\ \text{NPV}_{\text{avp}} &= (B_p - C_p) / P \\ \text{BCRp} &= B_p / C_p \\ \text{DPp} &= \text{Min } j \text{ such that } B_j > C_j \end{aligned}$$

Where:

$$\begin{aligned} \text{NPV}_p &= \text{Net present value to all participants} \\ \text{NPV}_{\text{avp}} &= \text{Net present value to the average participant} \\ \text{BCRp} &= \text{Benefit-cost ratio to participants} \\ \text{DPp} &= \text{Discounted payback in years} \\ B_p &= \text{NPV of benefit to participants} \\ C_p &= \text{NPV of costs to participants} \\ B_j &= \text{Cumulative benefits to participants in year } j \\ C_j &= \text{Cumulative costs to participants in year } j \\ P &= \text{Number of program participants} \\ J &= \text{First year in which cumulative benefits are cumulative costs.} \\ d &= \text{Interest rate (discount)} \end{aligned}$$

The Benefit (Bp) and Cost (Cp) terms are further defined as follows:

$$BP = \sum_{t=1}^N \frac{BR_t + TC_t + INC_t}{(1+d)^{t-1}} + \sum_{t=1}^N \frac{AB_{at} + PA_{at}}{(1+d)^{t-1}}$$

$$C = \sum_{t=1}^N \frac{PC_t + BI_t}{(1+d)^{t-1}}$$

Where:

$$\begin{aligned} \text{BR}_t &= \text{Bill reductions in year } t \\ \text{BI}_t &= \text{Bill increases in year } t \end{aligned}$$

TC _t	=	Tax credits in year t
INC _t	=	Incentives paid to the participant by the sponsoring utility in year t ³
PC _t	=	Participant costs in year t to include: • Initial capital costs, including sales tax⁴ • Ongoing operation and maintenance costs include fuel cost • Removal costs, less salvage value • Value of the customer's time in arranging for installation, if significant
PAC _{at}	=	Participant avoided costs in year t for alternate fuel devices (costs of devices not chosen)
Abat	=	Avoided bill from alternate fuel in year t

The first summation in the B_p equation should be used for conservation and load management programs. For fuel substitution programs, both the first and second summations should be used for B_p.

Note that in most cases, the customer bill impact terms (BR_t, BI_t, and AB_{at}) are further determined by costing period to reflect load impacts and/or rate schedules, which vary substantially by time of day and season. The formulas for these variables are as follows:

$$BR_t = \sum_{i=1}^I (\Delta EG_{it} \times AC : E_{it} \times K_{it}) + \sum_{i=1}^I (\Delta DG_{it} \times AC : D_{it} \times K_{it}) + OBR_t$$

AB_{at} = (Use BR_t formula, but with rates and costing periods appropriate for the alternate fuel utility)

$$BI_t = \sum_{i=1}^I (\Delta EG_{it} \times AC : E_{it} \times (K_{it} - 1)) + \sum_{i=1}^I (\Delta DG_{it} \times AC : D_{it} \times (K_{it} - 1)) + OBI_t$$

Where:

ΔEG _{it}	=	Reduction in gross energy use in costing period i in year t
ΔDG _{it}	=	Reduction in gross billing demand in costing period i in year t
AC:E _{it}	=	Rate charged for energy in costing period i in year t

³ Some difference of opinion exists as to what should be called an incentive. The term can be interpreted broadly to include almost anything. Direct rebates, interest payment subsidies, and even energy audits can be called incentives. Operationally, it is necessary to restrict the term to include only dollar benefits such as rebates or rate incentives (monthly bill credits). Information and services such as audits are not considered incentives for the purposes of these tests. If the incentive is to offset a specific participant cost, as in a rebate-type incentive, the full customer cost (before the rebate must be included in the PC_t term

⁴ If money is borrowed by the customer to cover this cost, it may not be necessary to calculate the annual mortgage and discount this amount if the present worth of the mortgage payments equals the initial cost. This occurs when the discount rate used is equal to the interest rate of the mortgage. If the two rates differ (e.g., a loan offered by the utility), then the stream of mortgage payments should be discounted by the discount rate chosen.

$AC:D_{it}$	=	Rate charged for demand in costing period i in year t
K_{it}	=	1 when ΔEG_{it} or ΔDG_{it} is positive (a reduction) in costing period i in year t, and zero otherwise
OBR_t	=	Other bill reductions or avoided bill payments (e.g., customer charges, standby rates).
OBI_t	=	Other bill increases (i.e. customer charges, standby rates).
I	=	Number of periods of participant's participation

In load management programs such as TOU rates and air-conditioning cycling, there are often no direct customer hardware costs. However, attempts should be made to quantify indirect costs customers may incur that enable them to take advantage of TOU rates and similar programs.

If no customer hardware costs are expected or estimates of indirect costs and value of service are unavailable, it may not be possible to calculate the benefit-cost ratio and discounted payback period.

Chapter 3

The Ratepayer Impact Measure Test⁵

Definition

The Ratepayer Impact Measure (RIM) test measures what happens to customer bills or rates due to changes in utility revenues and operating costs caused by the program. Rates will go down if the change in revenues from the program is greater than the change in utility costs. Conversely, rates or bills will go up if revenues collected after program implementation are less than the total costs incurred by the utility in implementing the program. This test indicates the direction and magnitude of the expected change in customer bills or rate levels.

Benefits and Costs

The benefits calculated in the RIM test are the savings from avoided supply costs. These avoided costs include the reduction in transmission, distribution, generation, and capacity costs for periods when load has been reduced and the increase in revenues for any periods in which load has been increased. The avoided supply costs are a reduction in total costs or revenue requirements and are included for both fuels for a fuel substitution program. The increase in revenues are also included for both fuels for fuel substitution programs. Both the reductions in supply costs and the revenue increases should be calculated using net energy savings.

The costs for this test are the program costs incurred by the utility, *and/or other entities incurring costs and creating or administering the program*, the incentives paid to the participant, decreased revenues for any periods in which load has been decreased and increased supply costs for any periods when load has been increased. The utility program costs include initial and annual costs, such as the cost of equipment, operation and maintenance, installation, program administration, and customer dropout and removal of equipment (less salvage value). The decreases in revenues and the increases in the supply costs should be calculated for both fuels for fuel substitution programs using net savings.

How the Results can be Expressed

The results of this test can be presented in several forms: the lifecycle revenue impact (cents or dollars) per kWh, kW, therm, or customer; annual or first-year revenue impacts (cents or dollars per kWh, kW, therms, or customer); benefit-cost ratio; and net present value. The primary units of measurement are the lifecycle revenue impact, expressed as the change in rates (cents per kWh for electric energy, dollars per kW for electric capacity, cents per therm for natural gas) and the net present value. Secondary test results are the lifecycle revenue

⁵ The Ratepayer Impact Measure Test has previously been described under what was called the "Non-Participant Test." The Non-Participant Test has also been called the "Impact on Rate Levels Test."

impact per customer, first-year and annual revenue impacts, and the benefit-cost ratio. LRI_{RIM} values for programs affecting electricity and gas should be calculated for each fuel individually (cents per kWh or dollars per kW and cents per therm) and on a combined gas and electric basis (cents per customer).

The lifecycle revenue impact (LRI) is the one-time change in rates or the bill change over the life of the program needed to bring total revenues in line with revenue requirements over the life of the program. The rate increase or decrease is expected to be put into effect in the first year of the program. Any successive rate changes such as for cost escalation are made from there. The first-year revenue impact (FRI) is the change in rates in the first year of the program or the bill change needed to get total revenues to match revenue requirements only for that year. The annual revenue impact (ARI) is the series of differences between revenues and revenue requirements in each year of the program. This series shows the cumulative rate change or bill change in a year needed to match revenues to revenue requirements. Thus, the $ARIRIM$ for year six per kWh is the estimate of the difference between present rates and the rate that would be in effect in year six due to the program. For results expressed as lifecycle, annual, or first-year revenue impacts, negative results indicate favorable effects on the bills of ratepayers or reductions in rates. Positive test result values indicate adverse bill impacts or rate increases.

Net present value (NPV_{RIM}) gives the discounted dollar net benefit of the program from the perspective of rate levels or bills over some specified time period. A net present value above zero indicates that the program will benefit (lower) rates and bills.

The benefit-cost ratio (BCR RIM) is the ratio of the total benefits of a program to the total costs discounted over some specified time period. A benefit-cost ratio above one indicates that the program will lower rates and bills.

Strengths of the Ratepayer Impact Measure (RIM) Test

In contrast to most supply options, demand-side management programs cause a direct shift in revenues. Under many conditions, revenues lost from DSM programs have to be made up by ratepayers. The RIM test is the only test that reflects this revenue shift along with the other costs and benefits associated with the program.

An additional strength of the RIM test is that the test can be used for all demand-side management programs (conservation, load management, fuel substitution, and load building). This makes the RIM test particularly useful for comparing impacts among demand-side management options.

Some of the units of measurement for the RIM test are of greater value than others, depending upon the purpose or type of evaluation. The lifecycle revenue impact per customer is the most useful unit of measurement when comparing the merits of programs with highly variable scopes (e.g., funding levels) and when analyzing a wide range of programs that

include both electric and natural gas impacts. Benefit-cost ratios can also be very useful for program design evaluations to identify the most attractive programs or program elements.

If comparisons are being made between a program or group of conservation/load management programs and a specific resource project, lifecycle cost per unit of energy and annual and first-year net costs per unit of energy are the most useful way to express test results. Of course, this requires developing lifecycle, annual, and first-year revenue impact estimates for the supply-side project.

Weaknesses of the Ratepayer Impact Measure (RIM) Test

Results of the RIM test are probably less certain than those of other tests because the test is sensitive to the differences between long-term projections of marginal costs and long-term projections of rates, two cost streams that are difficult to quantify with certainty.

RIM test results are also sensitive to assumptions regarding the financing of program costs. Sensitivity analyses and interactive analyses that capture feedback effects between system changes, rate design options, and alternative means of financing generation and non-generation options can help overcome these limitations. However, these types of analyses may be difficult to implement.

An additional caution must be exercised in using the RIM test to evaluate a fuel substitution program with multiple end use efficiency options. For example, under conditions where marginal costs are less than average costs, a program that promotes an inefficient appliance may give a more favorable test result than a program that promotes an efficient appliance. Though the results of the RIM test accurately reflect rate impacts, the implications for long-term conservation efforts need to be considered.

Formulae: The formulae for the lifecycle revenue impact (LRI RIM)' net present value (NPV RIM), benefit-cost ratio (BCR RIM)' the first-year revenue impacts and annual revenue impacts are presented below:

$$\begin{aligned} \text{LRIRIM} &= (\text{CRIM} - \text{BRIM}) / E \\ \text{FRIRIM} &= (\text{CRIM} - \text{BRIM}) / E && \text{for } t = I \\ \text{ARIRIM}_t &= \text{FRIRIM} && \text{for } t = I \\ &= (\text{CRIM}_t - \text{BRIM}_t) / E_t && \text{for } t=2, \dots, N \\ \text{NPVRIM} &= \text{BRIM} - \text{CRIM} \end{aligned}$$

$$\text{BCRRIM} = \text{BRIM} / \text{CRIM} \text{ where:}$$

LRIRIM = Lifecycle revenue impact of the program per unit of energy (kWh or therm) or demand (kW) (the one-time change in rates) or per customer (the change in customer bills over the life of the program). (Note: An appropriate choice of kWh, therm, kW, and customer should be made)

FRIRIM = First-year revenue impact of the program per unit of energy, demand, or per customer.

ARIRIM = Stream of cumulative annual revenue impacts of the program per unit of energy, demand, or per customer. (Note: The terms in the ARI formula are not discounted; thus they are the nominal cumulative revenue impacts. Discounted cumulative revenue impacts may be calculated and submitted if they are indicated as such. Note also that the sum of the discounted stream of cumulative revenue impacts does not equal the LRI RIM')

NPVRIM = Net present value levels

BCRRIM = Benefit-cost ratio for rate levels

BRIM = Benefits to rate levels or customer bills

CRIM = Costs to rate levels or customer bills

E = Discounted stream of system energy sales (kWh or therms) or demand sales (kW) or first-year customers. (See Appendix D for a description of the derivation and use of this term in the LRIRIM test.)

The B_{RIM} and C_{RIM} terms are further defined as follows:

$$B_{RIM} = \sum_{t=1}^N \frac{UAC_t + RG_t}{(1+d)^{t-1}} + \sum_{t=1}^N \frac{UAC_{at}}{(1+d)^{t-1}}$$

$$C_{RIM} = \sum_{t=1}^N \frac{UIC_t + RL_t + PRC_t + INC_t}{(1+d)^{t-1}} + \sum_{t=1}^N \frac{RL_{at}}{(1+d)^{t-1}}$$

$$E = \sum_{t=1}^N \frac{E_t}{(1+d)^{t-1}}$$

Where:

UAC_t = Utility avoided supply costs in year t
 UIC_t = Utility increased supply costs in year t
 RG_t = Revenue gain from increased sales in year t
 RL_t = Revenue loss from reduced sales in year t
 PRC_t = Program Administrator program costs in year t
 E_t = System sales in kWh, kW or therms in year t or first year customers
 UAC_{at} = Utility avoided supply costs for the alternate fuel in year t
 Rlat = Revenue loss from avoided bill payments for alternate fuel in year t (i.e., device not chosen in a fuel substitution program)

For fuel substitution programs, the first term in the B RIM and C RIM equations represents the sponsoring utility (electric or gas), and the second term represents the alternate utility. The RIM test should be calculated separately for electric and gas and combined electric and gas.

The utility avoided cost terms (UAC_t , UIC_t , and UAC_{at}) are further determined by costing period to reflect time-variant costs of supply:

$$UCA_t = \sum_{i=1}^I (\Delta EN_{it} \times MC : E_{it} \times K_{it}) + \sum_{i=1}^I (\Delta DN_{it} \times MC : D_{it} \times K_{it})$$

UAC_{at} = (Use UAC_t formula, but with marginal costs and costing periods appropriate for the alternate fuel utility.)

$$UIC_t = \sum_{i=1}^I (\Delta EN_{it} \times MC : E_{it} \times (K_{it} - 1)) + \sum_{i=1}^I (\Delta DN_{it} \times MC : D \times (K_{it} - 1))$$

Where:

[Only terms not previously defined are included here.]

- ΔEN_{it} = Reduction in net energy use in costing period i in year t
- ΔDN_{it} = Reduction in net demand in costing period i in year t
- $MC:E_{it}$ = Marginal cost of energy in costing period i in year t
- $MC:D_{it}$ = Marginal cost of demand in costing period i in year t

The revenue impact terms (RG_t , RL_t , and RL_{at}) are parallel to the bill impact terms in the Participant Test. The terms are calculated exactly the same way with the exception that the net impacts are used rather than gross impacts. If a net-to-gross ratio is used to differentiate gross savings from net savings, the revenue terms and the participant's bill terms will be related as follows:

- RG_t = BIt * (net-to-gross ratio)
- RL_t = BR_t * (net-to-gross ratio)
- RL_{at} = $Abat$ * (net-to-gross ratio)

Chapter 4

Total Resource Cost Test⁶

Definition

The Total Resource Cost Test measures the net costs of a demand-side management program as a resource option based on the total costs of the program, including both the participants' and the utility's costs.

The test is applicable to conservation, load management, and fuel substitution programs. For fuel substitution programs, the test measures the net effect of the impacts from the fuel not chosen versus the impacts from the fuel that is chosen as a result of the program. TRC test results for fuel substitution programs should be viewed as a measure of the economic efficiency implications of the total energy supply system (gas and electric).

A variant on the TRC test is the Societal Test. The Societal Test differs from the TRC test in that it includes the effects of externalities (e.g., environmental, national security), excludes tax credit benefits, and uses a different (societal) discount rate.

Benefits and Costs: This test represents the combination of the effects of a program on both the customers participating and those not participating in a program. In a sense, it is the summation of the benefit and cost terms in the Participant and the Ratepayer Impact Measure tests, where the revenue (bill) change and the incentive terms intuitively cancel (except for the differences in net and gross savings).

The benefits calculated in the Total Resource Cost Test are the avoided supply costs, the reduction in transmission, distribution, generation, and capacity costs valued at marginal cost for the periods when there is a load reduction. The avoided supply costs should be calculated using net program savings, savings net of changes in energy use that would have happened in the absence of the program. For fuel substitution programs, benefits include the avoided device costs and avoided supply costs for the energy, using equipment not chosen by the program participant.

The costs in this test are the program costs paid by both the utility and the participants plus the increase in supply costs for the periods in which load is increased. Thus all equipment costs, installation, operation and maintenance, cost of removal (less salvage value), and administration costs, no matter who pays for them, are included in this test. Any tax credits are considered a reduction to costs in this test. For fuel substitution programs, the costs also include the increase in supply costs for the utility providing the fuel that is chosen as a result of the program.

⁶ This test was previously called the All Ratepayers Test

How the Results Can be Expressed

The results of the Total Resource Cost Test can be expressed in several forms: as a net present value, a benefit-cost ratio, or as a levelized cost. The net present value is the primary unit of measurement for this test. Secondary means of expressing TRC test results are a benefit-cost ratio and levelized costs. The Societal Test expressed in terms of net present value, a benefit-cost ratio, or levelized costs is also considered a secondary means of expressing results. Levelized costs as a unit of measurement are inapplicable for fuel substitution programs, since these programs represent the net change of alternative fuels which are measured in different physical units (e.g., kWh or therms). Levelized costs are also not applicable for load building programs.

Net present value (NPVTRC) is the discounted value of the net benefits to this test over a specified period of time. NPVTRC is a measure of the change in the total resource costs due to the program. A net present value above zero indicates that the program is a less expensive resource than the supply option upon which the marginal costs are based.

The benefit-cost ratio (BCRTRC) is the ratio of the discounted total benefits of the program to the discounted total costs over some specified time period. It gives an indication of the rate of return of this program to the utility and its ratepayers. A benefit-cost ratio above one indicates that the program is beneficial to the utility and its ratepayers on a total resource cost basis.

The levelized cost is a measure of the total costs of the program in a form that is sometimes used to estimate costs of utility-owned supply additions. It presents the total costs of the program to the utility and its ratepayers on a per kilowatt, per kilowatt hour, or per therm basis levelized over the life of the program.

The Societal Test is structurally similar to the Total Resource Cost Test. It goes beyond the TRC test in that it attempts to quantify the change in the total resource costs to society as a whole rather than to only the service territory (the utility and its ratepayers). In taking society's perspective, the Societal Test utilizes essentially the same input variables as the TRC Test, but they are defined with a broader societal point of view. More specifically, the Societal Test differs from the TRC Test in at least one of five ways. First, the Societal Test may use higher marginal costs than the TRC test if a utility faces marginal costs that are lower than other utilities in the state or than its out-of-state suppliers. Marginal costs used in the Societal Test would reflect the cost to society of the more expensive alternative resources. Second, tax credits are treated as a transfer payment in the Societal Test, and thus are left out. Third, in the case of capital expenditures, interest payments are considered a transfer payment since society actually expends the resources in the first year. Therefore, capital costs enter the calculations in the year in which they occur. Fourth, a societal discount rate should be used⁷. Finally, Marginal costs used in the Societal Test would also contain externality costs of power generation not captured by the market system. An illustrative and

⁷ Many economists have pointed out that use of a market discount rate in social cost-benefit analysis undervalues the interests of future generations. Yet if a market discount rate is not used, comparisons with alternative investments are difficult to make.

by no means exhaustive list of ‘externalities and their components’ is given below (Refer to the Limitations section for elaboration.) These values are also referred to as ‘adders’ designed to capture or internalize such externalities. The list of potential adders would include for example:

1. The benefit of avoided environmental damage: The CPUC policy specifies two ‘adders’ to internalize environmental externalities, one for electricity use and one for natural gas use. Both are statewide average values. These adders are intended to help distinguish between cost-effective and non cost-effective energy-efficiency programs. They apply to an average supply mix and would not be useful in distinguishing among competing supply options. The CPUC electricity environmental adder is intended to account for the environmental damage from air pollutant emissions from power plants. The CPUC-adopted adder is intended to cover the human and material damage from sulfur oxides (SOX), nitrogen oxides (NOX), volatile organic compounds (VOC, sometimes called reactive organic gases or ROG), particulate matter at or below 10 micron diameter (PM10), and carbon. The adder for natural gas is intended to account for air pollutant emissions from the direct combustion of the gas. In the CPUC policy guidance, the adders are included in the tabulation of the benefits of energy efficiency programs. They represent reduced environmental damage from displaced electricity generation and avoided gas combustion. The environmental damage is the result of the net change in pollutant emissions in the air basins, or regions, in which there is an impact. This change is the result of direct changes in powerplant or natural gas combustion emission resulting from the efficiency measures, and changes in emissions from other sources, that result from those direct changes in emissions.
2. The benefit of avoided transmission and distribution costs – energy efficiency measures that reduce the growth in peak demand would decrease the required rate of expansion to the transmission and distribution network, eliminating costs of constructing and maintaining new or upgraded lines.
3. The benefit of avoided generation costs – energy efficiency measures reduce consumption and hence avoid the need for generation. This would include avoided energy costs, capacity costs and T&D line
4. The benefit of increased system reliability: The reductions in demand and peak loads from customers opting for self generation, provide reliability benefits to the distribution system in the forms of:
 - a. Avoided costs of supply disruptions
 - b. Benefits to the economy of damage and control costs avoided by customers and industries in the digital economy that need greater than 99.9 level of reliable electricity service from the central grid
 - c. Marginally decreased System Operator’s costs to maintain a percentage reserve of electricity supply above the instantaneous demand
 - d. Benefits to customers and the public of avoiding blackouts.

5. Non-energy benefits: Non-energy benefits might include a range of program-specific benefits such as saved water in energy-efficient washing machines or self generation units, reduced waste streams from an energy-efficient industrial process, etc.
6. Non-energy benefits for low income programs: The low income programs are social programs which have a separate list of benefits included in what is known as the 'low income public purpose test'. This test and the sepcific benefits associated with this test are outside the scope of this manual.
7. Benefits of fuel diversity include considerations of the risks of supply disruption, the effects of price volatility, and the avoided costs of risk exposure and risk management.

Strengths of the Total Resource Cost Test

The primary strength of the Total Resource Cost (TRC) test is its scope. The test includes total costs (participant plus program administrator) and also has the potential for capturing total benefits (avoided supply costs plus, in the case of the societal test variation, externalities). To the extent supply-side project evaluations also include total costs of generation and/or transmission, the TRC test provides a useful basis for comparing demand- and supply-side options.

Since this test treats incentives paid to participants and revenue shifts as transfer payments (from all ratepayers to participants through increased revenue requirements), the test results are unaffected by the uncertainties of projected average rates, thus reducing the uncertainty of the test results. Average rates and assumptions associated with how other options are financed (analogous to the issue of incentives for DSM programs) are also excluded from most supply-side cost determinations, again making the TRC test useful for comparing demand-side and supply-side options.

Weakness of the Total Resource Cost Test

The treatment of revenue shifts and incentive payments as transfer payments, identified previously as a strength, can also be considered a weakness of the TRC test. While it is true that most supply-side cost analyses do not include such financial issues, it can be argued that DSM programs should include these effects since, in contrast to most supply options, DSM programs do result in lost revenues.

In addition, the costs of the DSM "resource" in the TRC test are based on the total costs of the program, including costs incurred by the participant. Supply-side resource options are typically based only on the costs incurred by the power suppliers.

Finally, the TRC test cannot be applied meaningfully to load building programs, thereby limiting the ability to use this test to compare the full range of demand-side management options.

Formulas

The formulas for the net present value (NPV_{TRC})' the benefit-cost ratio (BCR_{TRC} and levelized costs are presented below:

$$\begin{aligned} \text{NPVTRC} &= \text{BTRC} - \text{CTRC} \\ \text{BCRTRC} &= \text{BTRC} / \text{CTRC} \\ \text{LCTRC} &= \text{LCRC} / \text{IMP} \end{aligned}$$

Where:

NPVTRC = Net present value of total costs of the resource
 BCRTRC = Benefit-cost ratio of total costs of the resource
 LCTRC = Levelized cost per unit of the total cost of the resource (cents per kWh for conservation programs; dollars per kW for load management programs)
 BTRC = Benefits of the program
 CTRC = Costs of the program
 LCRC = Total resource costs used for levelizing
 IMP = Total discounted load impacts of the program
 PCN = Net Participant Costs

The B_{TRC} C_{TRC} LCRC, and IMP terms are further defined as follows:

$$\text{BTRC} = \sum_{t=1}^N \frac{\text{UAC}_t + \text{TC}_t}{(1+d)^{t-1}} + \sum_{t=1}^N \frac{\text{UAC}_{at} + \text{PAC}_{at}}{(1+d)^{t-1}}$$

$$\text{CTRC} = \sum_{t=1}^N \frac{\text{PRC}_t + \text{PCN}_t + \text{UIC}_t}{(1+d)^{t-1}}$$

$$\text{LCRC} = \sum_{t=1}^N \frac{\text{PRC}_t + \text{PCN}_t - \text{TC}_t}{(1+d)^{t-1}}$$

$$\text{IMP} = \sum_{t=1}^n \left[\frac{\left(\sum_{i=1}^n \Delta \text{EN}_{it} \right) \text{ or } \left(\Delta \text{DN}_{it} \text{ where } I = \text{peak period} \right)}{(1+d)^{t-1}} \right]$$

[All terms have been defined in previous chapters.]

The first summation in the BTRC equation should be used for conservation and load management programs. For fuel substitution programs, both the first and second summations should be used.

Chapter 5

Program Administrator Cost Test

Definition

The Program Administrator Cost Test measures the net costs of a demand-side management program as a resource option based on the costs incurred by the program administrator (including incentive costs) and excluding any net costs incurred by the participant. The benefits are similar to the TRC benefits. Costs are defined more narrowly.

Benefits and Costs

The benefits for the Program Administrator Cost Test are the avoided supply costs of energy and demand, the reduction in transmission, distribution, generation, and capacity valued at marginal costs for the periods when there is a load reduction. The avoided supply costs should be calculated using net program savings, savings net of changes in energy use that would have happened in the absence of the program. For fuel substitution programs, benefits include the avoided supply costs for the energy-using equipment not chosen by the program participant only in the case of a combination utility where the utility provides both fuels.

The costs for the Program Administrator Cost Test are the program costs incurred by the administrator, the incentives paid to the customers, and the increased supply costs for the periods in which load is increased. Administrator program costs include initial and annual costs, such as the cost of utility equipment, operation and maintenance, installation, program administration, and customer dropout and removal of equipment (less salvage value). For fuel substitution programs, costs include the increased supply costs for the energy-using equipment chosen by the program participant only in the case of a combination utility, as above.

In this test, revenue shifts are viewed as a transfer payment between participants and all ratepayers. Though a shift in revenue affects rates, it does not affect revenue requirements, which are defined as the difference between the net marginal energy and capacity costs avoided and program costs. Thus, if $NPV_{pa} > 0$ and $NPV_{RIM} < 0$, the administrator's overall total costs will decrease, although rates may increase because the sales base over which revenue requirements are spread has decreased.

How the Results Can be Expressed

The results of this test can be expressed either as a net present value, benefit-cost ratio, or levelized costs. The net present value is the primary test, and the benefit-cost ratio and levelized cost are the secondary tests.

Net present value (NPV_{pa}) is the benefit of the program minus the administrator's costs, discounted over some specified period of time. A net present value above zero indicates that this demand-side program would decrease costs to the administrator and the utility.

The benefit-cost ratio (BCR_{pa}) is the ratio of the total discounted benefits of a program to the total discounted costs for a specified time period. A benefit-cost ratio above one indicates that the program would benefit the combined administrator and utility's total cost situation.

The levelized cost is a measure of the costs of the program to the administrator in a form that is sometimes used to estimate costs of utility-owned supply additions. It presents the costs of the program to the administrator and the utility on per kilowatt, per kilowatt-hour, or per therm basis levelized over the life of the program.

Strengths of the Program Administrator Cost Test

As with the Total Resource Cost test, the Program Administrator Cost test treats revenue shifts as transfer payments, meaning that test results are not complicated by the uncertainties associated with long-term rate projections and associated rate design assumptions. In contrast to the Total Resource Cost test, the Program Administrator Test includes only the portion of the participant's equipment costs that is paid for by the administrator in the form of an incentive. Therefore, for purposes of comparison, costs in the Program Administrator Cost Test are defined similarly to those supply-side projects which also do not include direct customer costs.

Weaknesses of the Program Administrator Cost Test

By defining device costs exclusively in terms of costs incurred by the administrator, the Program Administrator Cost test results reflect only a portion of the full costs of the resource.

The Program Administrator Cost Test shares two limitations noted previously for the Total Resource Cost test: (1) by treating revenue shifts as transfer payments, the rate impacts are not captured, and (2) the test cannot be used to evaluate load building programs.

Formulas

The formulas for the net present value, the benefit-cost ratio and levelized cost are presented below:

$$\begin{aligned} \text{NPV}_{pa} &= B_{pa} - C_{pa} \\ \text{BCR}_{pa} &= B_{pa}/C_{pa} \\ \text{LC}_{pa} &= \text{LC}_{pa}/\text{IMP} \end{aligned}$$

Where:

NPV _{pa}	Net present value of Program Administrator costs
BCR _{pa}	Benefit-cost ratio of Program Administrator costs

LCpa	Levelized cost per unit of Program Administrator cost of the resource
Bpa	Benefits of the program
Cpa	Costs of the program
LCpc	Total Program Administrator costs used for levelizing

$$B_{pa} = \sum_{t=1}^N \frac{UAC_t}{(1+d)^{t-1}} + \sum_{t=1}^N \frac{UAC_{at}}{(1+d)^{t-1}}$$

$$C_{pa} = \sum_{t=1}^N \frac{PRC_t + INC_t + UIC_t}{(1+d)^{t-1}}$$

$$LCpc = \sum_{t=1}^N \frac{PRC_t + INC_t}{(1+d)^{t-1}}$$

[All variables are defined in previous chapters.]

The first summation in the Bpa equation should be used for conservation and load management programs. For fuel substitution programs, both the first and second summations should be used.

Appendix A

Inputs to Equations and Documentation

A comprehensive review of procedures and sources for developing inputs is beyond the scope of this manual. It would also be inappropriate to attempt a complete standardization of techniques and procedures for developing inputs for such parameters as load impacts, marginal costs, or average rates. Nevertheless, a series of guidelines can help to establish acceptable procedures and improve the chances of obtaining reasonable levels of consistent and meaningful cost-effectiveness results. The following "rules" should be viewed as appropriate guidelines for developing the primary inputs for the cost-effectiveness equations contained in this manual:

1. In the past, Marginal costs for electricity were based on production cost model simulations that clearly identify key assumptions and characteristics of the existing generation system as well as the timing and nature of any generation additions and/or power purchase agreements in the future. With a deregulated market for wholesale electricity, marginal costs for electric generation energy should be based on forecast market prices, which are derived from recent transactions in California energy markets. Such transactions could include spot market purchases as well as longer term bilateral contracts and the marginal costs should be estimated based on components for energy as well as demand and/or capacity costs as is typical for these contracts.
2. In the case of submittals in conjunction with a utility rate proceeding, average rates used in DSM program cost-effectiveness evaluations should be based on proposed rates. Otherwise, average rates should be based on current rate schedules. Evaluations based on alternative rate designs are encouraged.
3. Time-differentiated inputs for electric marginal energy and capacity costs, average energy rates, and demand charges, and electric load impacts should be used for (a) load management programs, (b) any conservation program that involves a financial incentive to the customer, and (c) any Fuel Substitution or Load Building program. Costing periods used should include, at a minimum, summer and winter, on-, and off-peak; further disaggregation is encouraged.
4. When program participation includes customers with different rate schedules, the average rate inputs should represent an average weighted by the estimated mix of participation or impacts. For General Rate Case proceedings it is likely that each major rate class within each program will be considered as program elements requiring separate cost-effectiveness analyses for each measure and each rate class within each program.

5. Program administration cost estimates used in program cost-effectiveness analyses should exclude costs associated with the measurement and evaluation of program impacts unless the costs are a necessary component to administer the program.
6. For DSM programs or program elements that reduce electricity and natural gas consumption, costs and benefits from both fuels should be included.
7. The development and treatment of load impact estimates should distinguish between gross (i.e., impacts expected from the installation of a particular device, measure, appliance) and net (impacts adjusted to account for what would have happened anyway, and therefore not attributable to the program). Load impacts for the Participants test should be based on gross, whereas for all other tests the use of net is appropriate. Gross and net program impact considerations should be applied to all types of demand-side management programs, although in some instances there may be no difference between gross and net.
8. The use of sensitivity analysis, i.e. the calculation of cost-effectiveness test results using alternative input assumptions, is encouraged, particularly for the following programs: new programs, programs for which authorization to substantially change direction is being sought (e.g., termination, significant expansion), major programs which show marginal cost-effectiveness and/or particular sensitivity to highly uncertain input(s).

The use of many of these guidelines is illustrated with examples of program cost effectiveness contained in Appendix B.

Appendix B

Summary of Equations and Glossary of Symbols

Basic Equations

Participant Test

$$\begin{aligned} \text{NPVP} &= \text{BP} - \text{CP} \\ \text{NPV}_{\text{avp}} &= (\text{BP} - \text{CP}) / P \\ \text{BCRP} &= \text{BP} / \text{CP} \\ \text{DPP} &= \min j \text{ such that } B_j > C_j \end{aligned}$$

Ratepayer Impact Measure Test

$$\begin{aligned} \text{LRIRIM} &= (\text{CRIM} - \text{BRIM}) / E \\ \text{FRIRIM} &= (\text{CRIM} - \text{BRIM}) / E && \text{for } t = 1 \\ \text{ARIRIM}_t &= \text{FRIRIM} && \text{for } t = 1 \\ &= (\text{CRIM}_t - \text{BRIM}_t) / E_t && \text{for } t=2, \dots, N \\ \text{NPVRIM} &= \text{BRIM} - \text{CRIM} \\ \text{BCRRIM} &= \text{BRIM} / \text{CRIM} \end{aligned}$$

Total Resource Cost Test

$$\begin{aligned} \text{NPVTRC} &= \text{BTRC} - \text{CTRC} \\ \text{BCRTRC} &= \text{BTRC} / \text{CTRC} \\ \text{LCTRC} &= \text{LCRC} / \text{IMP} \end{aligned}$$

Program Administrator Cost Test

$$\begin{aligned} \text{NPV}_{\text{pa}} &= B_{\text{pa}} - C_{\text{pa}} \\ \text{BCR}_{\text{pa}} &= B_{\text{pa}} / C_{\text{pa}} \\ \text{LC}_{\text{pa}} &= \text{LC}_{\text{pa}} / \text{IMP} \end{aligned}$$

Benefits and Costs

Participant Test

$$Bp = \sum_{t=1}^N \frac{BR_t + TC_t + INC_t}{(1+d)^{t-1}} + \sum_{t=1}^N \frac{AB_{at} + PAC_{at}}{(1+d)^{t-1}}$$

$$Cp = \sum_{t=1}^N \frac{PC_t + BI_t}{(1+d)^{t-1}}$$

Ratepayer Impact Measure Test

$$B_{RIM} = \sum_{t=1}^N \frac{UAC_t + RG_t}{(1+d)^{t-1}} + \sum_{t=1}^N \frac{UAC_{at}}{(1+d)^{t-1}}$$

$$C_{RIM} = \sum_{t=1}^N \frac{UIC_t + RL_t + PRC_t + INC_t}{(1+d)^{t-1}} + \sum_{t=1}^N \frac{RL_{at}}{(1+d)^{t-1}}$$

$$E = \sum_{t=1}^N \frac{E_t}{(1+d)^{t-1}}$$

Total Resource Cost Test

$$B_{TRC} = \sum_{t=1}^N \frac{UAC_t + TC_t}{(1+d)^{t-1}} + \sum_{t=1}^N \frac{UAC_{at} + PAC_{at}}{(1+d)^{t-1}}$$

$$C_{TRC} = \sum_{t=1}^N \frac{PRC_t + PCN_t + UIC_t}{(1+d)^{t-1}}$$

$$L_{TRC} = \sum_{t=1}^N \frac{PRC_t + PCN_t - TC_t}{(1+d)^{t-1}}$$

$$IMP = \frac{\sum_{t=1}^n \left[\left(\sum_{i=1}^n \Delta EN_{it} \right) \text{ or } (\Delta DN_{it} \text{ where } I = \text{peak period}) \right]}{(1+d)^{t-1}}$$

Program Administrator Cost Test

$$B_{pa} = \sum_{t=1}^N \frac{UAC_t}{(1+d)^{t-1}} + \sum_{t=1}^N \frac{UAC_{at}}{(1+d)^{t-1}}$$

$$C_{pa} = \sum_{t=1}^N \frac{PRC_t + INC_t + UIC_t}{(1+d)^{t-1}}$$

$$LCPA = \sum_{t=1}^N \frac{PRC_t + INC_t}{(1+d)^{t-1}}$$

Glossary of Symbols

Abat	=	Avoided bill reductions on bill from alternate fuel in year t
AC:Dit	=	Rate charged for demand in costing period i in year t
AC:Eit	=	Rate charged for energy in costing period i in year t
ARIRIM	=	Stream of cumulative annual revenue impacts of the program per unit of energy, demand, or per customer. Note that the terms in the ARI formula are not discounted, thus they are the nominal cumulative revenue impacts. Discounted cumulative revenue impacts may be calculated and submitted if they are indicated as such. Note also that the sum of the discounted stream of cumulative revenue impacts does not equal the LRIRIM*
BCRp	=	Benefit-cost ratio to participants
BCRRIM	=	Benefit-cost ratio for rate levels
BCRTRC	=	Benefit-cost ratio of total costs of the resource
BCRpa	=	Benefit-cost ratio of program administrator and utility costs
Bl _t	=	Bill increases in year t
B _j	=	Cumulative benefits to participants in year j
B _p	=	Benefit to participants
BRIM	=	Benefits to rate levels or customer bills
BR _t	=	Bill reductions in year t
BTRC	=	Benefits of the program
B _{pa}	=	Benefits of the program
C _j	=	Cumulative costs to participants in year i

Cp	= Costs to participants
CRIM	= Costs to rate levels or customer bills
CTRC	= Costs of the program
Cpa	= Costs of the program
D	= discount rate
ΔD_{git}	= Reduction in gross billing demand in costing period i in year t
ΔD_{nit}	= Reduction in net demand in costing period i in year t
DPp	= Discounted payback in years
E	= Discounted stream of system energy sales-(kWh or therms) or demand sales (kW) or first-year customers
ΔE_{git}	= Reduction in gross energy use in costing period i in year t
ΔE_{nit}	= Reduction in net energy use in costing period i in year t
E_t	= System sales in kWh, kW or therms in year t or first year customers
FRIRIM	= First-year revenue impact of the program per unit of energy, demand, or per customer.
IMP	= Total discounted load impacts of the program
INCt	= Incentives paid to the participant by the sponsoring utility in year t First year in which cumulative benefits are > cumulative costs.
Kit	= 1 when ΔE_{Git} or ΔD_{Git} is positive (a reduction) in costing period i in year t, and zero otherwise
LCRC	= Total resource costs used for levelizing
LCTRC	= Levelized cost per unit of the total cost of the resource
LCPA	= Total Program Administrator costs used for levelizing
Lcpa	= Levelized cost per unit of program administrator cost of the resource
LRIRIM	= Lifecycle revenue impact of the program per unit of energy (kWh or therm) or demand (kW)-the one-time change in rates-or per customer-the change in customer bills over the life of the program.
MC:Dit	= Marginal cost of demand in costing period i in year t
MC:Eit	= Marginal cost of energy in costing period i in year t
NPVavp	= Net present value to the average participant
NPVP	= Net present value to all participants
NPVRIM	= Net present value levels
NPVTRC	= Net present value of total costs of the resource
NPVpa	= Net present value of program administrator costs
OBI _t	= Other bill increases (i.e., customer charges, standby rates)
OB _{Rt}	= Other bill reductions or avoided bill payments (e.g., customer charges, standby rates).
P	= Number of program participants
PACat	= Participant avoided costs in year t for alternate fuel devices

PCt	= Participant costs in year t to include: • Initial capital costs, including sales tax • Ongoing operation and maintenance costs • Removal costs, less salvage value • Value of the customer's time in arranging for installation, if significant
PRCt	= Program Administrator program costs in year t
PCN	= Net Participant Costs
RGt	= Revenue gain from increased sales in year t
RLat	= Revenue loss from avoided bill payments for alternate fuel in year t (i.e., device not chosen in a fuel substitution program)
RLt	= Revenue loss from reduced sales in year t
TCt	= Tax credits in year t
UACat	= Utility avoided supply costs for the alternate fuel in year t
UACt	= Utility avoided supply costs in year t
PA _t	= Program Administrator costs in year t
UICt	= Utility increased supply costs in year t

Appendix C.

Derivation of Rim Lifecycle Revenue Impact Formula

Most of the formulas in the manual are either self-explanatory or are explained in the text. This appendix provides additional explanation for a few specific areas where the algebra was considered to be too cumbersome to include in the text.

Rate Impact Measure

The Ratepayer Impact Measure lifecycle revenue impact test (LRIRIM) is assumed to be the one-time increase or decrease in rates that will re-equate the present valued stream of revenues and stream of revenue requirements over the life of the program.

Rates are designed to equate long-term revenues with long-term costs or revenue requirements. The implementation of a demand-side program can disrupt this equality by changing one of the assumptions upon which it is based: the sales forecast. Demand-side programs by definition change sales. This expected difference between the long-term revenues and revenue requirements is calculated in the NPVRIM. The amount which present valued revenues are below present valued revenue requirements equals NPVRIM.

The LRIRIM is the change in rates that creates a change in the revenue stream that, when present valued, equals the NPVRIM*. If the utility raises (or lowers) its rates in the base year by the amount of the LRIRIM, revenues over the term of the program will again equal revenue requirements. (The other assumed changes in rates, implied in the escalation of the rate values, are considered to remain in effect.)

Thus, the formula for the LRIRIM is derived from the following equality where the present value change in revenues due to the rate increase or decrease is set equal to the NPVRIM or the revenue change caused by the program.

$$-NPV_{RIM} = \sum_{t=1}^N \frac{LRI_{RIM} \times E_t}{(1+d)^{t-1}}$$

Since the LRI_{RIM} term does not have a time subscript, it can be removed from the summation, and the formula is then:

$$-NPV_{RIM} = LRI_{RIM} \times \sum_{t=1}^N \frac{E_t}{(1+d)^{t-1}}$$

Rearranging terms, we then get:

$$LRI_{RIM} = -NPV_{RIM} \bigg/ \sum_{t=1}^N \frac{E_t}{(1+d)^{t-1}}$$

Thus,

$$E = \sum_{t=1}^N \frac{E_t}{(1+d)^{t-1}}$$



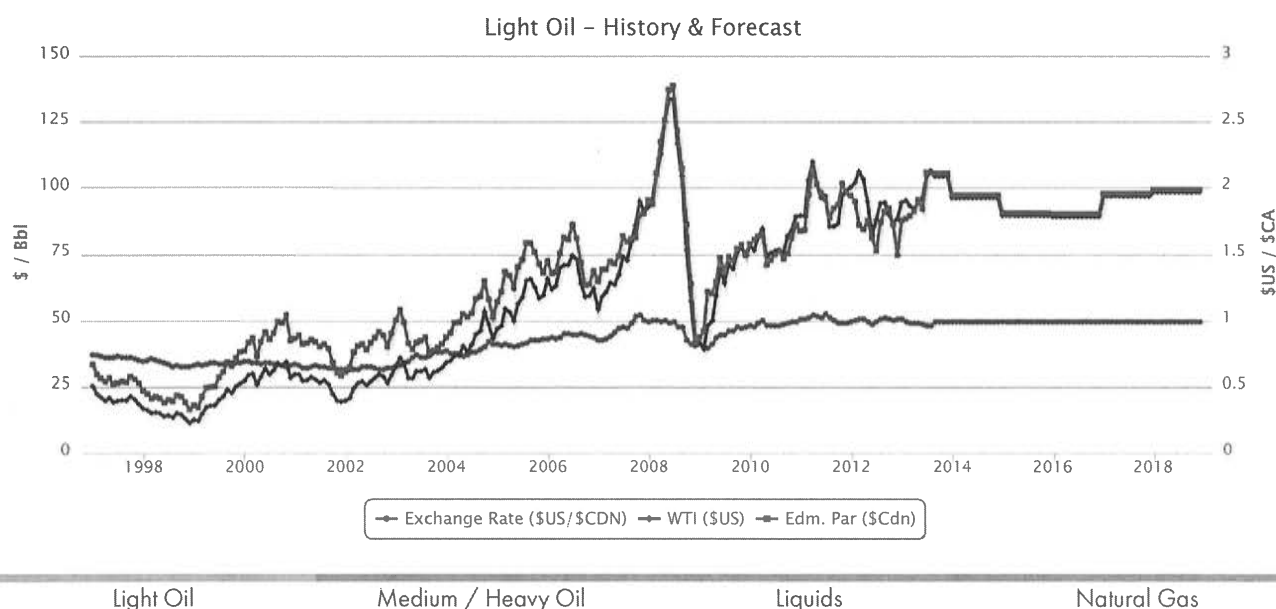
Forecasts

August 31, 2013

Every month, Sproule prepares product price and market forecast tables for the Canadian oil and gas industry. The forecasts are based on information obtained from various sources, including government agencies, industry publications, Canadian oil refiners, and natural gas marketers.

Download

-  [Escalated Forecast](#)
-  [Constant Forecast](#)



Disclaimer

The price tables linked on this page present the product prices and market forecasts for the Canadian oil and gas industry prepared by Sproule*. These prices and forecasts are based on information obtained from various sources, including government agencies, industry publications, Canadian oil refiners, and natural gas marketers, available up to and including the "as of" date in the table. The information was accepted as is, and Sproule accepts no responsibility for any inaccuracies within it. The forecasts presented here are based on an informed interpretation of the information and can be considered reasonable at the time they were published.

Users of this information are asked to recognize the high degree of uncertainty associated with forecasting oil and gas prices, and Sproule takes no responsibility for the application of these numbers by anyone other than the professionals of Sproule. Users of this information do so at their own risk, and Sproule and its officers, directors, consultants and employees shall not be held responsible or liable to any users for any reliance that may be placed on the information.

These forecasts are revised monthly. Revisions could be considerable. Sproule and its subsidiaries are not liable for any errors, omissions or inaccuracies in the numbers provided on this site.

Monthly Newsletter

Join our mailing list to get notified as soon as the newest forecast becomes available

SIGN UP NOW!

Forecast Archives

To see price forecasts for previous months from 2004 to present, please see our

PRICE ARCHIVES

**"Sproule" refers to any of the Sproule group of companies, including Sproule Associates Ltd., Sproule International Ltd., Sproule U.S. Ltd., Sproule Unconventional Ltd. and Sproule Project Management Ltd.*

Copyright © 2013 Sproule - Worldwide Petroleum Consultants
[home](#) | [contact info](#)

3 Results

3.1 Quantitative results

The results of the study are summarized in Table 6 through Table 8 below. They are broken down into four general life cycle stages:

1. Transport generator – from manufacturing location to Whitehorse
2. Produce fuel – includes treatment, refining and transport to facilities
3. Transport fuel –diesel/LNG to Whitehorse
4. Combust fuel – power generation

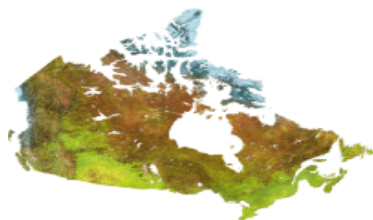
These tables provide the specific quantitative results for each system, by stage, for each parameter assessed.

Table 6. Diesel pathway results by life cycle stage

	GHGs (kgCO ₂ e/MWH)	NO _x (kg/MWH)	SO ₂ (kg/MWH)	PM (kg/MWH)	CO (kg/MWH)
Transport diesel generator	0.06	0.0	0.0	0.0	0.0
Produce diesel	254	0.4	0.4	0.3	0.3
Transport diesel	11	0.2	0.0	0.0	0.0
Combust diesel	619	14.1	0.0	0.8	1.0
TOTAL	884	14.7	0.4	1.2	1.3

Table 7. LNG pathway results by life cycle stage

	GHGs (kgCO ₂ e/MWH)	NO _x (kg/MWH)	SO ₂ (kg/MWH)	PM (kg/MWH)	CO (kg/MWH)
Transport natural gas generator	0.05	0.0	0.0	0.00	0.0
Produce LNG	176	0.2	1.5	0.01	0.1
Transport LNG	24	0.3	0.0	0.01	0.0
Combust natural gas	488	0.1	0.0	0.01	0.2
TOTAL	688	0.6	1.5	0.03	0.3



PRODUCTION OF CANADA'S LEADING MINERALS

PRODUCTION DES PRINCIPAUX MINÉRAUX DU CANADA

1 Shipments of Selected Minerals / Livraisons de certains minéraux

	Cement ¹ / Ciment ¹		Clay products / Produits d'argile		Gypsum / Gypse		Lime / Chaux	
	2010	2011	2010	2011	2010	2011	2010	2011
	tonnes		dollars		tonnes		tonnes	
January - Janvier	511 117	486 379	9 183 380 r	7 096 681	166 792 r	161 987	157 973	148 444
February - Février	519 525	472 297	9 382 363 r	6 671 842	282 692 r	228 513	157 077	134 036
March - Mars	715 436	667 093	15 265 849 r	10 534 476	276 580 r	224 665	163 938	169 400
April - Avril	945 599	785 218	15 857 745 r	11 809 876	208 541 r	188 293	149 548	156 973
May - Mai	1 138 308	1 088 927	15 346 365 r	12 284 329	322 904 r	144 167	161 787	179 978
June - Juin	1 181 335	1 193 781	13 688 226 r	15 666 689	280 585 r	233 897	161 655	168 284
July - Juillet	1 131 180	1 173 333	12 114 283 r	12 264 405	227 484 r	96 258	155 606	164 953
August - Août	1 261 375	1 281 793	14 289 627 r	14 133 604	262 321 r	211 837	158 398	165 431
September - Septembre	1 163 924	1 227 329	12 940 778 r	13 779 093	261 150 r	212 080	155 477	169 221
October - Octobre	1 250 231	1 229 587	13 088 250 r	11 463 971	221 409 r	132 238	146 521	176 619
November - Novembre	1 083 515	1 100 317	11 578 269 r	12 634 028	288 170 r	184 643	142 061 r	152 895
December - Décembre	621 014	701 540	6 171 412 r	7 052 603	247 647 r	215 128	152 953	150 770 r
Total	11 522 559	11 407 594	148 906 547	135 391 597	3 046 275	2 233 706	1 862 995	1 937 004

	Molybdenum / Molybdène		Potash / Potasse (K ₂ O)		Diamonds ² / Diamants ²		Uranium (U) / Uranium (U)	
	2010	2011	2010	2011	2010	2011	2010	2011
	kilograms - kilogrammes		tonnes		carats		kilograms - kilogrammes	
January - Janvier	722 618	652 500	714 416 r	987 491	785 424	812 056	899 502	862 793
February - Février	777 244	592 399	849 538 r	933 142	873 880	768 324	800 043	816 988
March - Mars	771 366	765 005	1 022 740 r	1 135 021	1 214 685	891 300	899 045	400 290
April - Avril	804 132	586 437	939 075 r	941 230	941 400	918 586	807 459	604 603
May - Mai	659 638	468 962	800 539 r	1 052 907	1 030 139	935 673	725 346	954 890
June - Juin	586 268	679 772	653 504 r	955 114	996 075	1 011 334	1 314 144	890 781
July - Juillet	787 728	573 244	547 074 r	715 687	1 132 879	971 828	404 346	140 241
August - Août	833 782	734 724	645 506 r	696 698	962 504	954 434	596 361	793 045
September - Septembre	606 611	930 557	740 845 r	960 845	1 025 419	916 732	830 715	950 599
October - Octobre	602 870	907 817	887 347 r	759 766	1 140 121	801 831	1 208 023	919 970
November - Novembre	637 682	821 234	918 988 r	827 031	814 959	913 556	820 754	607 551
December - Décembre	734 392	893 362	980 192 r	758 267	886 610	856 348	620 829	1 075 117
Total	8 524 330	8 606 014	9 699 764	10 723 200	11 804 095	10 752 001	9 926 566	9 016 868

r Revised figures

¹ Source: Statistics Canada

² Production

For definitions regarding data contained in this publication refer to Natural Resources Canada 26-202, 2010. Components may not add due to rounding.

r: chiffres révisés

¹ Source: Statistique Canada

² Production

Pour connaître les définitions de données contenues dans cette publication, veuillez consulter le n° du catalogue 26-202 de Ressources naturelles Canada, 2010. Les chiffres ayant été arrondis leur somme peut ne pas correspondre aux totaux indiqués.



Natural Resources
Canada

Ressources naturelles
Canada



2 Production of Nickel / Production de nickel

	Mine output ¹ / Production minière ¹		Primary production ² / Production de première fusion ²		Refined production / Production affiné		Domestic shipments of refined production / Livraisons intérieures de production affiné	
	2010	2011	2010	2011	2010	2011	2010	2011
	tonnes							
January - Janvier	8 206 r	16 454	12 065 r	15 858	6 638	12 848	124	351
February - Février	6 532 r	17 795	6 332 r	17 143	6 302	12 657	167	297
March - Mars	9 003 r	19 203	8 706 r	18 501	7 455	13 251	131	359
April - Avril	9 537 r	21 294	9 222 r	20 507	7 197	12 714	113	265
May - Mai	12 074 r	16 150	11 675 r	15 561	7 278	11 265	126	257
June - Juin	11 018 r	15 326	10 663 r	14 783	6 727	6 983	260	344
July - Juillet	13 287 r	19 282	12 870 r	18 565	6 205	11 995	183	235
August - Août	12 639 r	15 596	12 238 r	15 025	4 211	9 306	217	278
September - Septembre	18 750 r	17 662	18 100 r	17 007	11 392	10 785	254	242
October - Octobre	21 599 r	19 802	20 857 r	19 087	13 887	13 477	334	354
November - Novembre	18 483 r	22 039	17 848 r	21 235	13 199	14 636	362	234
December - Décembre	18 935 r	19 008	16 594 r	18 314	14 922	12 529	253	324
Total	160 063	219 612	157 170	211 587	105 413	142 445	2 524	3 539

¹ Metal content in concentrates produced.¹ Teneur en métal des concentrés produits.² Recoverable metal in concentrates shipped.² Métal récupérable contenu dans les concentrés livrés.

3 Production of Copper / Production de cuivre

	Mine output ¹ / Production minière ¹		Primary production ² / Production de première fusion ²		Refined production / Production affiné		Domestic shipments of refined production / Livraisons intérieures de production affiné	
	2010	2011	2010	2011	2010	2011	2010	2011
	tonnes							
January - Janvier	37 853 r	46 615	38 705 r	46 623	32 546	23 039	10 030	12 361
February - Février	35 383 r	43 891	37 921 r	42 248	30 084	21 928	9 583	12 022
March - Mars	39 417 r	49 042	40 162 r	48 388	33 628	24 713	11 857	14 382
April - Avril	35 261 r	47 820	33 185 r	47 989	31 623	21 532	12 031	11 250
May - Mai	41 108 r	46 840	38 058 r	39 294	26 417	21 604	14 342	12 810
June - Juin	39 184 r	42 546	37 210 r	42 466	23 585	18 295	14 774	13 150
July - Juillet	42 118 r	47 252	42 345 r	39 427	22 640	20 205	9 787	8 846
August - Août	46 004 r	50 738	45 506 r	60 111	23 767	24 231	12 618	11 562
September - Septembre	50 400 r	44 388	48 486 r	40 423	21 833	23 152	13 304	11 496
October - Octobre	52 341 r	48 680	50 066 r	48 782	22 119	25 696	11 738	12 763
November - Novembre	48 810 r	49 205	44 497 r	44 346	25 675	24 461	11 904	11 391
December - Décembre	54 293 r	49 106	51 741 r	48 548	25 703 r	24 904	8 812	7 650
Total	522 172	566 124	507 883	548 645	319 618	273 761	140 780	139 682

¹ Metal content in concentrates produced.¹ Teneur en métal des concentrés produits.² Recoverable metal in concentrates shipped.² Métal récupérable contenu dans les concentrés livrés.

4 Production, Shipments and Stocks of Producers of Salt / Production, expéditions et stocks des producteurs de sel

	Fine vacuum / Fin, traité sous vide		Mined rock salt / Sel extrait des mines		Salt content of brines / Teneur en sel des saumures		Total	
	2010	2011	2010	2011	2010	2011	2010	2011
tonnes								
Production								
January - Janvier	72 282	74 150	1 023 020	1 046 642	9 460	10 485	1 104 762	1 131 277
February - Février	63 234	65 224	613 036	884 499	7 947	8 922	684 217	958 645
March - Mars	72 120	67 541	742 016	805 512	9 015	10 188	823 151	883 241
April - Avril	73 420	78 333	684 350	893 514	10 005	8 726	767 775	980 573
May - Mai	76 161	75 376	805 910	1 023 507	15 825	13 352	897 896	1 112 235
June - Juin	73 498	67 839	580 972	1 013 216	16 073	15 229	670 543	1 096 284
July - Juillet	74 451	74 397	632 805	988 911	12 303	14 195	719 559	1 077 503
August - Août	78 772	69 380	642 615	724 075	13 004	13 989	734 391	807 444
September - Septembre	67 998	66 338	844 089	908 439	9 860	10 265	921 947	985 042
October - Octobre	77 438	69 902	975 469	1 086 861	10 824	10 704	1 063 731	1 167 467
November - Novembre	74 441	67 515	1 007 398	1 127 632	9 234	9 888	1 091 073	1 205 035
December - Décembre	63 264 r	58 860	984 608	1 127 021	12 838 r	10 002	1 060 710 r	1 195 883
Total	867 079	834 855	9 536 288	11 629 829	136 388	135 945	10 539 755	12 600 629

Shipped or used by producers - Quantité expédiées ou utilisées par les producteurs

January - Janvier	66 785	72 540	1 158 692	1 228 848	9 460	10 485	1 234 937 r	1 311 873
February - Février	59 931	64 711	528 176	945 361	7 947	8 922	596 054 r	1 018 994
March - Mars	76 667	86 020	484 966 r	550 637	9 015	10 188	570 648 r	646 845
April - Avril	75 336	73 005	518 507	751 884	10 005	8 726	603 848 r	833 615
May - Mai	74 936	75 130	852 654	1 030 590	15 825	13 352	943 415 r	1 119 072
June - Juin	77 418	70 922	593 781	1 199 916	16 073	15 229	687 272 r	1 286 067
July - Juillet	69 459	68 695	731 215	880 698	12 303	14 195	812 977 r	963 588
August - Août	77 607	69 220	648 610	811 351	13 004	13 989	739 221 r	894 560
September - Septembre	75 767	68 543	826 787	722 497	9 860	10 265	912 414 r	801 305
October - Octobre	71 920	64 657	856 560	1 120 940	10 824	10 704	939 304 r	1 196 301
November - Novembre	78 151	67 494	1 040 574	1 269 752	9 234	9 888	1 127 959 r	1 347 134
December - Décembre	68 522	60 778	1 028 726	1 264 278	12 838 r	10 002	1 110 086 r	1 335 058
Total	872 499	841 715	9 269 248	11 776 752	136 388	135 945	10 278 135	12 754 412

Stocks at end of month - Stocks à la fin du mois

January - Janvier	67 122	60 911	344 845	663 807	-	-	411 967	724 718
February - Février	70 425	61 424	429 705	602 945	-	-	500 130	664 369
March - Mars	65 878	42 945	686 755	857 820	-	-	752 633	900 765
April - Avril	63 962	48 273	852 598	999 450	-	-	916 560	1 047 723
May - Mai	65 187	48 519	805 854	992 367	-	-	871 041	1 040 886
June - Juin	61 267	45 436	793 045	805 667	-	-	854 312	851 103
July - Juillet	66 259	51 138	694 635	913 880	-	-	760 894	965 018
August - Août	67 424	51 298	688 640	826 604	-	-	756 064	877 902
September - Septembre	59 655	49 093	705 942	1 012 546	-	-	765 597	1 061 639
October - Octobre	65 173	54 338	824 851	978 467	-	-	890 024	1 032 805
November - Novembre	61 463	54 359	791 675	836 347	-	-	853 138	890 706
December - Décembre	56 205	52 441	747 557	699 090	-	-	803 762	751 531

5 Gold Production¹, by Region / Production¹ d'or, par région

	Atlantic Provinces / Provinces de l'Atlantique		Québec		Ontario		Prairie Provinces / Provinces des Prairies	
	2010	2011	2010	2011	2010	2011	2010	2011
	grams - grammes							
January - Janvier	32 989	30 939	2 230 752	1 589 578	4 039 493	3 764 989 r	523 546	470 326
February - Février	38 520	26 081	2 098 853	1 516 942	3 524 370	3 702 495 r	326 521	507 984
March - Mars	37 448	26 096	2 107 786	1 986 385	6 002 812	6 841 802 r	465 054	606 961
April - Avril	41 749	30 009	2 247 092	1 603 602	3 277 561	3 388 013 r	514 282	464 346
May - Mai	35 602	58 251	2 103 686	1 981 882	3 567 978	3 521 866 r	470 398	545 120
June - Juin	34 218	51 381	2 168 499	2 471 578	4 296 283	5 916 990 r	476 270	603 542
July - Juillet	29 686	57 869	1 975 949	2 017 192	3 784 328	3 846 402 r	482 819	523 457
August - Août	22 127	59 174	1 795 803	2 168 493	4 696 576	3 712 273 r	594 873	541 570
September - Septembre	28 085	49 658	2 454 481	3 125 865	5 630 316	5 256 211 r	501 413	685 500
October - Octobre	26 185	43 040	2 068 382	2 288 706	4 248 321	3 879 108 r	528 849	684 348
November - Novembre	26 486	48 159	1 898 638	2 246 178	5 497 792	4 283 544 r	507 042	555 487
December - Décembre	28 878	67 229 r	2 176 597	2 948 090 r	6 188 760	4 548 212 r	491 253	648 743
Total	381 975	547 887	25 326 519	25 944 491	54 754 592	52 661 903	5 882 320	6 837 384

	British Columbia Colombie-Britannique		Yukon		Northwest Territories - Nunavut - Territoires du Nord-Ouest		Canada	
	2010	2011	2010	2011	2010	2011	2010	2011
	grams - grammes							
January - Janvier	485 253	399 376	60 409	65 321	0	640 078	7 372 442	6 960 607 r
February - Février	404 521	293 669	101 325	22 209	0	640 078	6 494 110	6 709 458 r
March - Mars	468 077	197 321	150 024	57 606	779 544	640 078	10 010 744	10 356 249 r
April - Avril	403 775	180 545	26 523	5 191	779 544	615 600	7 290 527	6 287 306 r
May - Mai	446 259	148 778	304	9 166	779 544	615 600	7 403 772	6 880 664 r
June - Juin	500 105	170 903	123 074	42 274 r	779 544	615 600	8 377 993	9 872 268 r
July - Juillet	452 579	125 075	166 496	134 921	779 544	810 152	7 671 402	7 515 068 r
August - Août	508 543	184 843	224 205	153 891	779 544	810 152	8 621 672	7 630 396 r
September - Septembre	463 412	153 963	559 373	232 212 r	779 544	810 152	10 416 624	10 313 562 r
October - Octobre	586 554	254 964	864 557	1 115 506	779 542	808 690	9 102 390	9 074 362 r
November - Novembre	422 389	209 482	61 280	90 933	779 542	811 801	9 193 171	8 245 584 r
December - Décembre	462 412	157 536	64 943	74 875	779 541	808 690	10 192 384	9 253 375 r
Total	5 603 879	2 476 456	2 402 514	2 004 104	7 795 433	8 626 674	102 147 230	99 098 899

¹ Recoverable gold in bullion and concentrates shipped.¹ Or récupérable contenu dans les lingots et les concentrés livrés.

6 Production of Gold / Production d'or

	Mine output ¹ / Production minière ¹		Production ² from Gold quartz mines and placer deposits / Production ² de mines de quartz aurifère et gisements alluvionnaires d'or		Production ² from base metal mines / Production ² de mines de métaux communs		Total production ² Production ² totale	
	2010	2011	2010	2011	2010	2011	2010	2011
	grams - grammes							
January - Janvier	7 864 905	7 684 886 r	7 125 710	6 292 747 r	739 195	667 860	7 372 442	6 960 607 r
February - Février	6 379 322	6 640 773 r	5 767 326	6 082 980 r	611 996	626 477	6 494 110	6 709 458 r
March - Mars	9 580 240	9 672 452 r	8 891 212	9 609 915 r	689 028	746 334	10 010 744	10 356 249 r
April - Avril	7 776 043	6 861 514 r	7 061 151	5 665 046 r	714 892	622 259	7 290 527	6 287 306 r
May - Mai	7 895 765	7 926 205 r	7 189 707	6 321 614 r	706 058	559 050	7 403 772	6 880 664 r
June - Juin	7 946 572	8 969 789 r	7 163 720	9 214 766 r	782 852	657 502	8 377 993	9 872 268 r
July - Juillet	7 940 870	8 020 650 r	7 279 245	6 704 465 r	661 625	810 604	7 671 402	7 515 068 r
August - Août	9 111 536	8 403 810 r	8 242 014	6 816 274 r	869 522	814 123	8 621 672	7 630 396 r
September - Septembre	9 358 127	9 507 482 r	8 619 143	9 627 478 r	738 984	686 084	10 416 624	10 313 562 r
October - Octobre	10 133 999	9 670 784 r	9 238 939	8 257 201 r	895 060	817 161	9 102 390	9 074 362 r
November - Novembre	9 244 487	8 455 958 r	8 438 875	7 502 720 r	805 612	742 864	9 193 171	8 245 584 r
December - Décembre	9 461 130	8 565 084 r	8 596 678	8 565 107 r	864 452	688 268	10 192 384	9 253 375 r
Total	102 692 996	100 379 387	93 613 720	90 660 313	9 079 276	8 438 586	102 147 230	99 098 899

¹ Gold content in bullion and concentrates produced.¹ Teneur en or des lingots et concentrés produits.² Recoverable gold in bullion and concentrates shipped.² Or récupérable contenu dans les lingots et les concentrés livrés.

7 Production of Lead / Production de plomb

	Mine output ¹ / Production minière ¹		Primary production ² / Production de première fusion ²		Refined production (primary) ³ / Production affiné (primaire) ³		Secondary production / Production de seconde fusion	
	2010	2011	2010	2011	2010	2011	2010	2011
	tonnes							
January - Janvier	5 865	5 502	6 248	4 126	10 276	7 902	13 425	14 652
February - Février	5 572	5 943	4 264	3 438	8 928	8 514	12 638	13 560
March - Mars	6 415	5 037	3 161	6 251	10 859	11 143	15 097	14 408
April - Avril	6 436	4 988	5 797	4 482	10 468	9 226	12 992	14 922
May - Mai	5 909	5 551	3 314	5 494	10 525	10 273	13 712	15 423
June - Juin	5 857	5 458	10 748	5 542	9 984	8 956	14 689	15 261
July - Juillet	4 622	5 061	3 921	4 201	9 705	8 511	13 556	11 604
August - Août	3 665	4 065	4 755	3 339	3 366	7 464	9 521	14 529
September - Septembre	5 616	3 291	4 701	6 032	9 236	9 522	14 404	13 241
October - Octobre	5 321	4 064	5 876	2 422	6 934	10 126	16 158	14 345
November - Novembre	4 612	4 999	3 230	3 376	4 864	9 829	14 871	13 764
December - Décembre	4 954	5 485	6 381	6 094	10 692	11 063	15 981	14 351 r
Total	64 844	59 445	62 397	54 797	105 836	112 531	167 042	170 059

¹ Metal content in concentrates produced.² Recoverable metal in concentrates shipped.³ Refined pig lead plus lead content of antimonial lead produced from domestic and foreign ores.¹ Teneur en métal des concentrés produits.² Métal récupérable contenu dans les concentrés livrés.³ Plomb en saumon affiné plus la teneur en plomb de plomb antimoné produit à partir de minerais canadiens et étrangers.

8 Production of Zinc / Production de zinc

	Mine output ¹ / Production minière ¹		Primary production ² / Production de première fusion ²		Refined production / Production affiné		Domestic shipments of refined production / Livraisons intérieures de production affiné	
	2010	2011	2010	2011	2010	2011	2010	2011
	tonnes							
January - Janvier	58 085 r	53 183	52 851 r	39 947	66 489	55 945	10 697	12 058
February - Février	51 326 r	52 646	39 032 r	42 012	58 616	48 911	13 222	10 868
March - Mars	61 771 r	56 714	63 284 r	63 487	67 118	56 722	11 974	12 457
April - Avril	62 764 r	56 344	53 686 r	45 193	65 387	54 912	13 432	11 727
May - Mai	57 119 r	57 920	52 177 r	53 608	56 668	55 829	12 870	13 110
June - Juin	58 608 r	53 534	59 830 r	66 405	50 822	54 786	14 288	14 124
July - Juillet	49 393 r	47 369	47 499 r	42 923	54 800	54 880	11 366	11 585
August - Août	50 386 r	49 258	46 607 r	41 879	52 052	50 563	13 461	10 844
September - Septembre	50 092 r	43 634	52 168 r	52 171	54 280	57 663	12 919	10 364
October - Octobre	53 236 r	46 666	43 450 r	40 158	55 937	54 295	12 685	11 568
November - Novembre	48 134 r	45 854	42 292 r	38 235	52 865	56 455	10 273	13 366
December - Décembre	48 151 r	48 455	56 691 r	54 514	56 188	61 190	10 088	12 150
Total	649 065	611 577	609 567	580 534	691 221	662 151	147 276	144 220

¹ Metal content in concentrates produced.² Recoverable metal in concentrates shipped.¹ Teneur en métal des concentrés produits.² Métal récupérable contenu dans les concentrés livrés.

9 Producers' Shipments of Iron Ore / Expéditions de minerai de fer des producteurs

	To Canadian consumers / Destiné au marché intérieur		Exported / Destiné à l'exportation		Total shipments / Livraisons totales		Stocks at end of month ¹ / Stocks à la fin du mois ¹	
	2010	2011	2010	2011	2010	2011	2010	2011
dryweight - tonnes - poids sec								
Newfoundland and Labrador - Terre-Neuve et Labrador								
January - Janvier	-	-	1 298 585	694 615	1 298 585	694 615	2 241 482 r	1 986 089
February - Février	-	-	1 125 410	1 216 903	1 125 410	1 216 903	2 223 686 r	1 493 171
March - Mars	-	24 042	1 390 815	1 295 976	1 390 815	1 320 018	2 282 011 r	1 353 186
April - Avril	75 599	51 516	1 424 124	1 021 426	1 499 723	1 072 942	2 529 926 r	1 209 238
May - Mai	25 666	129 245	1 810 342	1 324 193	1 836 008	1 453 438	2 214 473 r	1 560 156
June - Juin	97 332	25 415	1 871 141	1 133 690	1 968 473	1 159 105	1 677 097 r	1 668 323
July - Juillet	107 593	100 809	1 523 479	1 190 981	1 631 072	1 291 790	1 327 723 r	1 814 818
August - Août	103 520	52 397	1 250 416	1 989 052	1 353 936	2 041 449	1 598 711 r	1 724 072
September - Septembre	135 533	73 577	994 666	1 573 964	1 130 199	1 647 541	1 915 536 r	1 523 858
October - Octobre	50 379	102 228	1 082 661 r	1 283 591	1 133 040 r	1 385 819	2 443 616 r	1 643 896
November - Novembre	76 699	52 520	2 184 009 r	1 160 824	2 260 708 r	1 213 344	1 953 919 r	1 856 610
December - Décembre	991 558 r	104 405	1 456 421 r	1 235 995	2 447 979 r	1 340 400	795 373 r	1 863 084
Total	1 663 879	716 154	17 412 069	15 121 210	19 075 948	15 837 364		
Québec								
January - Janvier	138 437 r	138 020	993 223	923 723	1 131 660 r	1 061 743	1 235 507 r	2 352 707
February - Février	-	60 137	842 551 r	829 537	842 551 r	889 674	1 579 854 r	2 653 487
March - Mars	102 831 r	-	1 012 978 r	1 602 206	1 115 809 r	1 602 206	1 794 684 r	2 715 117
April - Avril	189 328	251 719	1 293 883 r	1 123 662	1 483 211 r	1 375 381	1 691 050 r	3 075 933
May - Mai	317 776 r	211 266	1 068 425	1 730 428	1 386 201 r	1 941 694	1 802 470 r	2 774 957
June - Juin	474 357 r	366 586	1 284 503	1 282 534	1 758 860 r	1 649 120	1 733 432 r	2 801 446
July - Juillet	228 947	315 984	1 087 875 r	1 616 108	1 316 822 r	1 932 092	2 023 548 r	2 700 725
August - Août	315 253 r	311 567	1 077 070	1 413 094	1 392 323 r	1 724 661	2 398 721 r	2 876 836
September - Septembre	236 723	413 154	1 567 528 r	1 196 643	1 804 251 r	1 609 797	2 113 779 r	2 847 661
October - Octobre	238 443	175 798	1 109 502	1 280 973	1 347 945	1 456 771	2 370 533 r	3 153 208
November - Novembre	290 747	343 652	1 703 535	1 991 479	1 994 282	2 335 131	1 738 010 r	2 433 178
December - Décembre	211 926	326 041	1 222 794 r	1 872 310	1 434 720 r	2 198 351	1 850 537 r	1 823 487
Total	2 744 768	2 913 924	14 263 867	16 862 697	17 008 634	19 776 621		
British Columbia - Colombie-Britannique								
January - Janvier	2 957	4 128	-	-	2 957	4 128	-	-
February - Février	3 079	2 265	-	-	3 079	2 265	-	-
March - Mars	5 997	4 100	-	-	5 997	4 100	-	-
April - Avril	6 885	5 585	-	-	6 885	5 585	-	-
May - Mai	10 033 r	7 089	-	-	10 033 r	7 089	-	-
June - Juin	10 058	8 109	-	-	10 058	8 109	-	-
July - Juillet	11 973	4 178	-	-	11 973	4 178	-	-
August - Août	12 316	6 342	-	-	12 316	6 342	-	-
September - Septembre	7 728	5 912	-	-	7 728	5 912	-	-
October - Octobre	6 758	6 131	-	-	6 758	6 131	-	-
November - Novembre	9 723	4 664	-	-	9 723	4 664	-	-
December - Décembre	5 138	1 121	-	-	5 807 r	1 121	-	-
Total	92 646	59 623	-	-	93 314	59 623		
Canada								
January - Janvier	141 394 r	142 148	2 291 808	1 618 338	2 433 202 r	1 760 486	3 476 989	4 338 796
February - Février	3 079	62 402	1 967 961 r	2 046 440	1 971 040 r	2 108 842	3 803 540	4 146 658
March - Mars	108 828 r	28 142	2 403 793 r	2 898 182	2 512 621 r	2 926 324	4 076 695	4 068 303
April - Avril	271 812	308 820	2 718 007 r	2 145 088	2 989 819 r	2 453 908	4 220 976	4 285 171
May - Mai	353 475 r	347 600	2 878 767	3 054 621	3 232 242 r	3 402 221	4 016 943	4 335 113
June - Juin	581 747 r	400 110	3 155 644	2 416 224	3 737 391 r	2 816 334	3 410 529	4 469 769
July - Juillet	348 513	420 971	2 611 354 r	2 807 089	2 959 867 r	3 228 060	3 351 271	4 515 543
August - Août	431 089 r	370 306	2 327 486	3 402 146	2 758 575 r	3 772 452	3 997 432	4 600 908
September - Septembre	379 984	492 643	2 562 194 r	2 770 607	2 942 178 r	3 263 250	4 029 315	4 371 519
October - Octobre	295 580	284 157	2 192 163 r	2 564 564	2 487 743 r	2 848 721	4 814 149	4 797 104
November - Novembre	377 169	400 836	3 887 544 r	3 152 303	4 264 713 r	3 553 139	3 691 929	4 289 788
December - Décembre	1 208 622 r	431 567	2 679 884 r	3 108 305	3 888 506 r	3 539 872	2 645 910	3 686 571
Total	4 501 292	3 689 701	31 676 605	31 983 907	36 177 897	35 673 608		

¹ Excludes stocks of crude ore¹ Exclut les stocks de minerai brut

10 Production of Silver / Production d'argent

	Mine output ¹ / Production minière ¹		Primary production ² / Production de première fusion ²	
	2010	2011	2010	2011
	kilograms - kilogrammes			
January - Janvier	55 105 r	47 907	50 329 r	41 763
February - Février	46 163 r	50 162	44 414 r	44 327
March - Mars	58 162 r	54 832	52 954 r	55 862
April - Avril	50 038 r	51 256	43 458 r	48 410
May - Mai	48 393 r	48 909	43 394 r	41 067
June - Juin	49 841 r	45 727	57 919 r	40 976
July - Juillet	45 245 r	48 879	44 312 r	44 646
August - Août	47 171 r	49 941	50 220 r	48 734
September - Septembre	48 175 r	42 586	45 009 r	43 197
October - Octobre	45 992 r	44 306	43 461 r	40 717
November - Novembre	48 108 r	44 118	46 016 r	37 813
December - Décembre	49 089 r	43 710 r	48 242 r	45 394 r
Total	591 482	572 333	569 729	532 907

¹ Metal content in concentrates produced.¹ Teneur en métal des concentrés produits.² Recoverable metal in concentrates shipped.² Métal récupérable contenu dans les concentrés livrés.

11 Production of Fuels / Production des combustibles

	Coal / Charbon		Natural gas ¹ / Gaz naturel ¹		Crude oil and equivalent ¹ / Pétrole brut et hydrocarbures équivalents ¹	
	2010	2011	2010	2011	2010	2011
	tonnes		millions m ³		000 m ³	
January - Janvier	5 830 000	5 254 000	13 921	14 261	12 945	14 634
February - Février	5 633 000	4 948 000	12 000	12 703	12 701	13 401
March - Mars	6 012 000	5 746 000	12 391	12 962	13 625	14 751
April - Avril	5 731 000	5 727 000	11 445	11 520	13 554	14 240
May - Mai	5 536 000	5 482 000	11 725	11 293	14 276	13 222
June - Juin	5 328 000	5 236 000	10 783	11 245	13 930	13 261
July - Juillet	5 698 000	5 544 000	11 545	11 319	14 374	14 985
August - Août	5 621 000	5 874 000	11 894	11 498	14 466	15 717
September - Septembre	5 448 000	5 650 000	11 108	10 965	13 327	14 799
October - Octobre	5 728 000	5 945 000	11 676	11 696	13 984	15 565
November - Novembre	5 656 000	5 582 000	11 840	11 744 r	14 802	15 013 r
December - Décembre	5 931 000 r	6 126 000	14 049	13 794	15 173	16 026
Total	68 152 000	67 114 000	144 378	144 999	167 157	175 613

¹ Marketable production¹ Production marchande

Source: Statistics Canada

Source: Statistique Canada

12 Production of Aluminum / Production d'aluminium

	Production molten metal / Production de métal en fusion		Metals added / Métaux d'addition		Total	
	2010	2011	2010	2011	2010	2011
	tonnes					
January - Janvier	256 878	253 261	4 112	3 779	260 990	257 040
February - Février	232 247	228 736	4 001	3 504	236 248	232 240
March - Mars	256 118	254 462	4 094	3 829	260 212	258 290
April - Avril	246 800	246 432	3 633	4 050	250 433	250 481
May - Mai	254 491	252 107	3 611	3 409	258 101	255 516
June - Juin	247 054	244 373	3 110	3 596	250 164	247 970
July - Juillet	243 254	251 266	4 147	3 843	247 401	255 108 r
August - Août	243 230	255 782	4 299	4 016	247 529	259 798
September - Septembre	237 755	245 609	3 805	3 498	241 560	249 107
October - Octobre	250 192	252 822	3 627	3 385	253 819	256 207
November - Novembre	243 393	246 816	3 446	3 560	246 839	250 377
December - Décembre	251 799	256 297 r	3 474	3 453	255 273	259 751
Total	2 963 210	2 987 964	45 359	43 922	3 008 569	3 031 886

Inquiries concerning this publication should be directed to:

Julie Simon, Statistical Production Coordinator
 Minerals and Mining Statistics Division
 Minerals and Metals Sector, Natural Resources Canada
 Telephone: (613) 947-6777 Fax: (613) 992-5565

Si vous avez des questions concernant cette publication,
veuillez communiquer avec:

Julie Simon, Coordinatrice de la production des statistiques
 Division de la statistique sur les minéraux et sur l'activité minière
 Secteur des minéraux et des métaux, Ressources naturelles Canada
 Téléphone: (613) 947-6777 Télécopieur: (613) 992-5565