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
- Jeorge 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

Table of Tables

Table of Figures

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 2004[1] (*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.

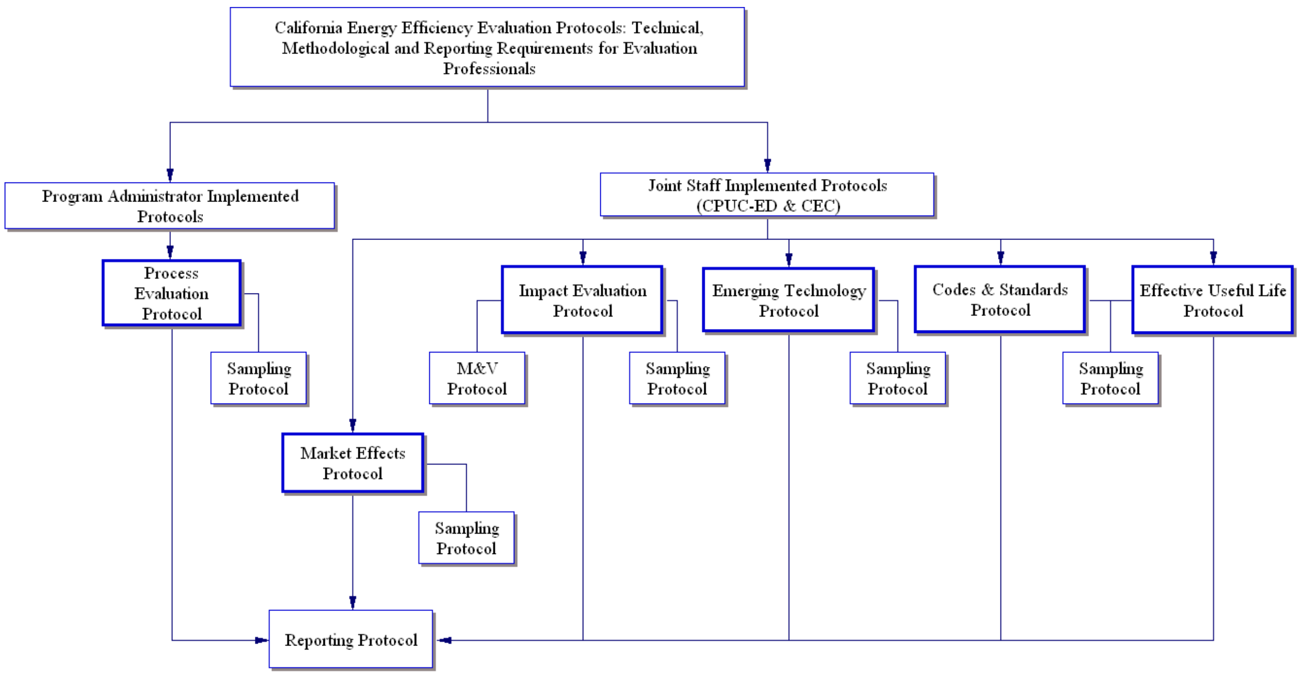

Figure 1. Operational Overview of How the Protocols Relate to Each Other

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 <u>and</u> 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 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.

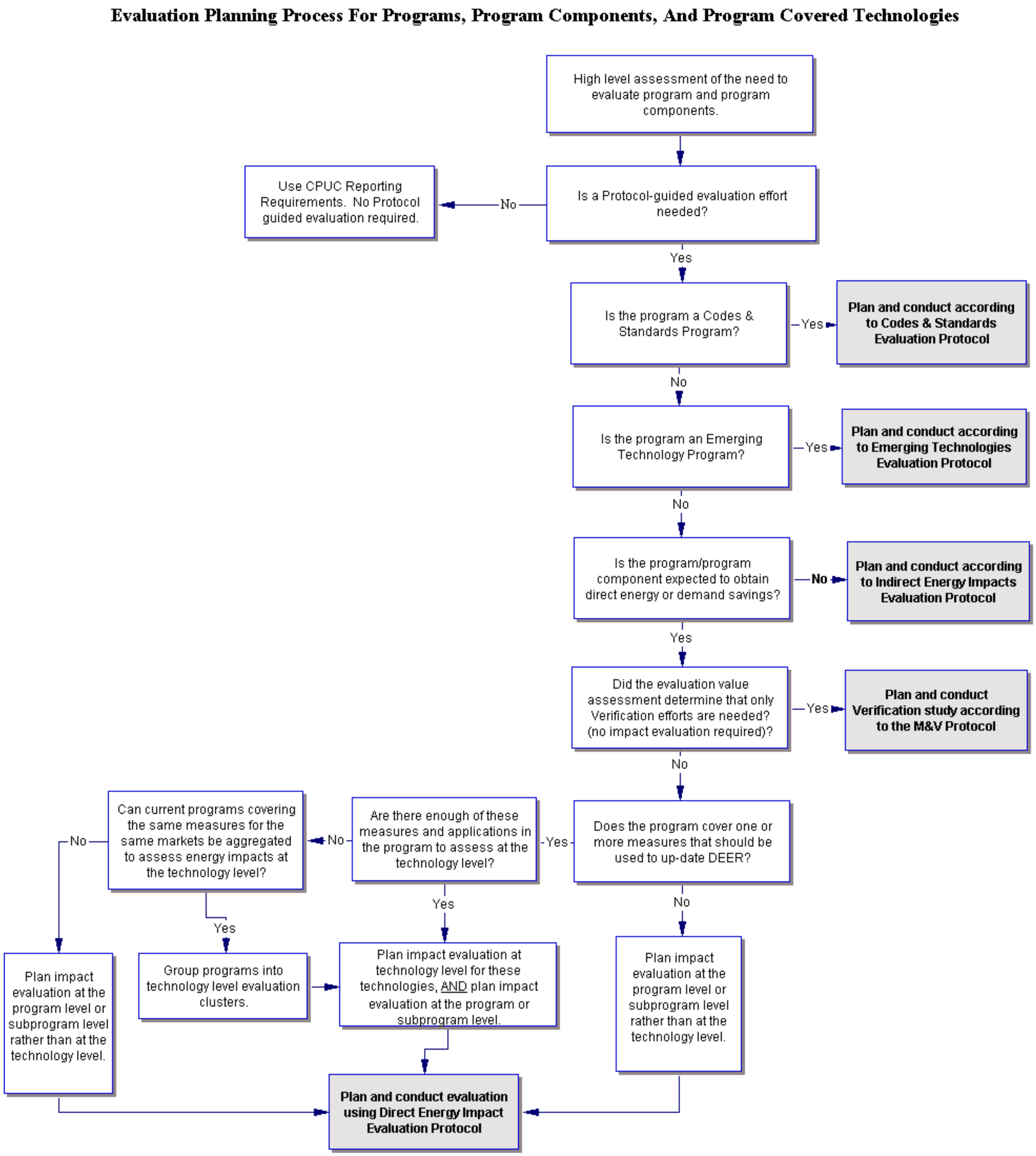

Figure 2. The Program Evaluation Planning Process for Programs, Program Components and Program-Covered Technologies

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).

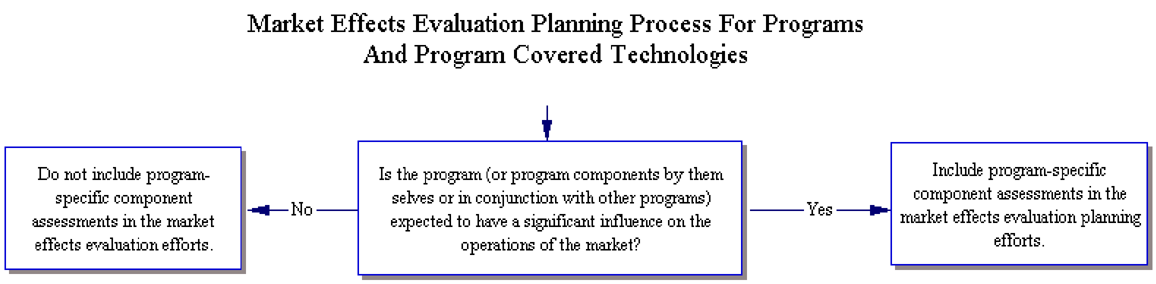

Figure 3. The Market Effects Evaluation Planning Process

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) option[6], 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 confidentially 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 indirect[7] 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 <u>Gross Energy Impact Protocol</u> 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.)

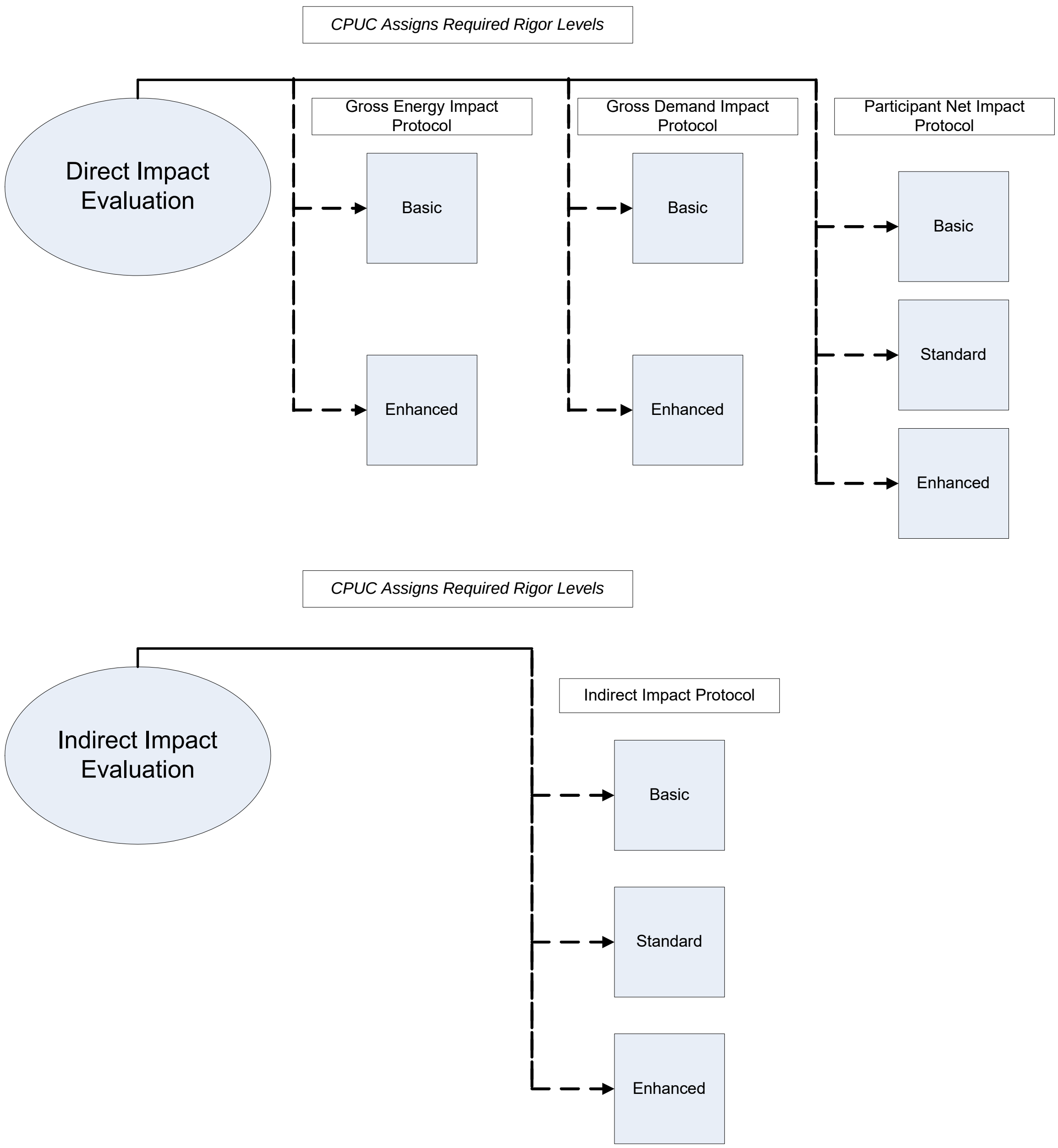

Figure 4. Required Protocols for Direct Impact and Indirect Impact Evaluations

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/demand[25] 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	1. Participant and non-participant analysis of utility consumption data that addresses the issue of self-selection. 2. 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. 3. Econometric or discrete choice[30] 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 triangulate[39] 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 *verify* 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.

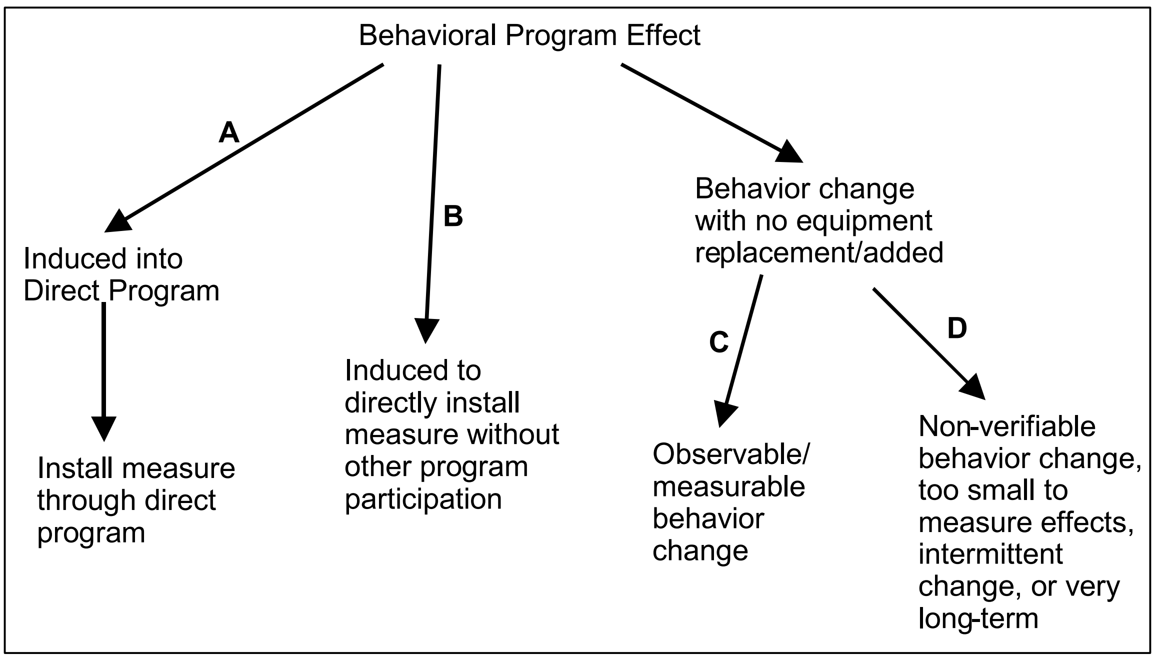

Figure 5. Potential Alternative Behavioral Impact Paths

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-efficiently 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 Managers[45] 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.

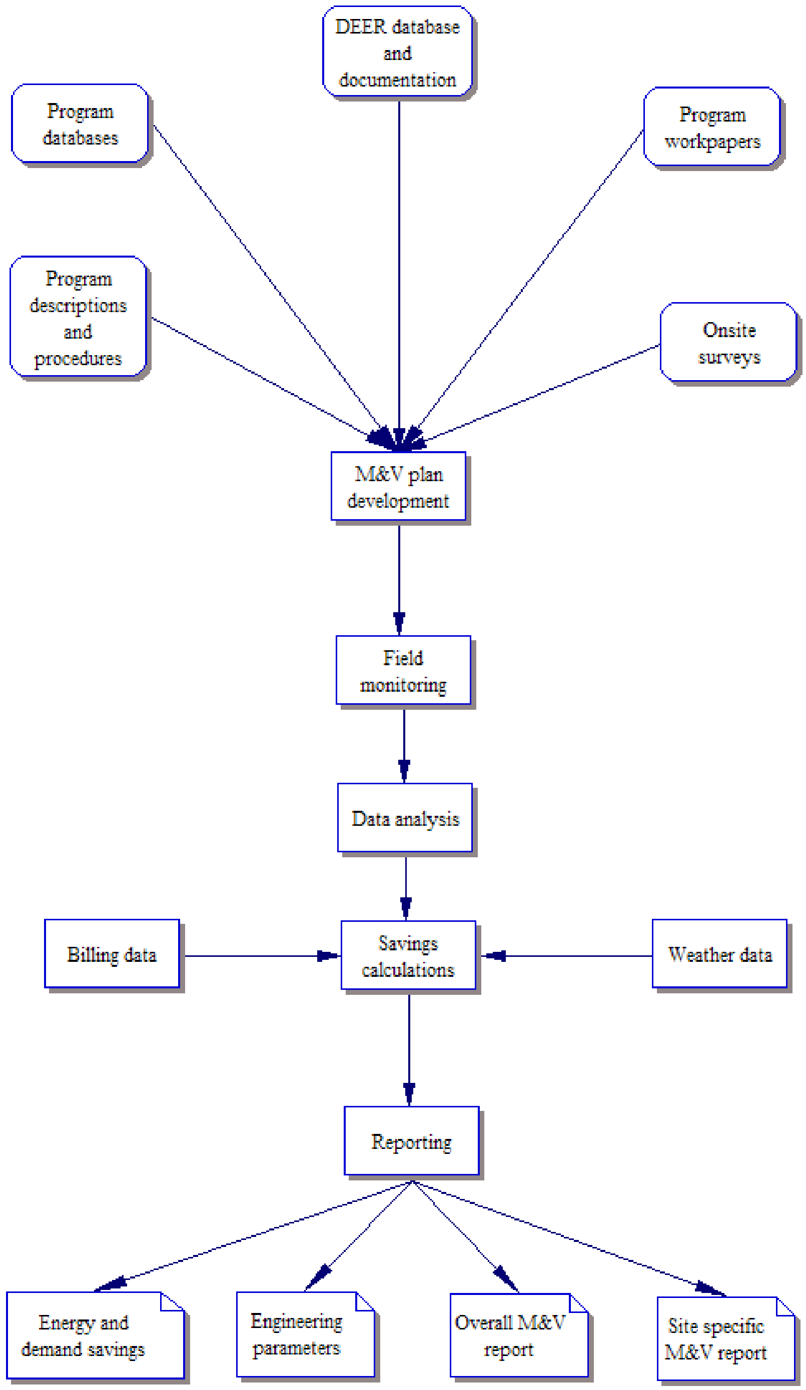

Figure 6. Measurement & Verification Information Flow Diagram

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 Isolation[55] 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	± 15%	± 5%
Hourly	± 30%	± 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 ET 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 failures[61] (technologies that do not perform as expected) given the inherent risks[62] 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 impacts[63].

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 error[64]. 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 assessments[65],
- 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 period[66],
- 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 rates[67] 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 projects[68] 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 panel[69], 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, surveys[71] 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 contactor 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 Prahl 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 Heschong 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 mitigated[72].

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 addresses[73]. 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 my 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 contactor 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 groups[74].

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 process[75]. 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 savings[76].

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 paper[77] 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 contactor 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 contactor 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 contactor 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 contactor 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 Delphi[78] 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 change[79].

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 contactor 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-compiling 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 or 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 equipment[81] 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.

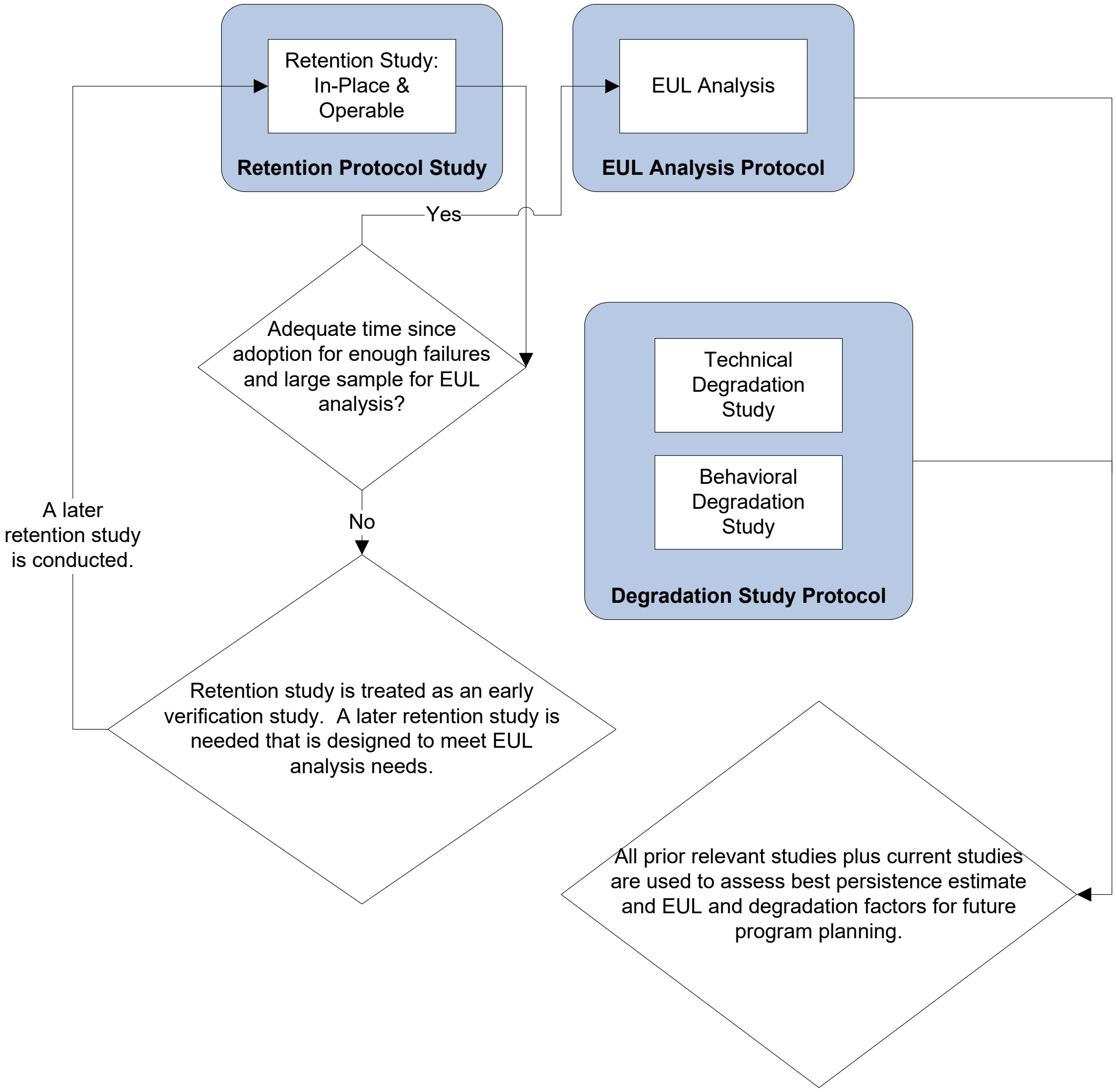

Figure 7. An Example of How Findings Across the Three Types of Studies Would Work Together for Persistence Evaluations

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).

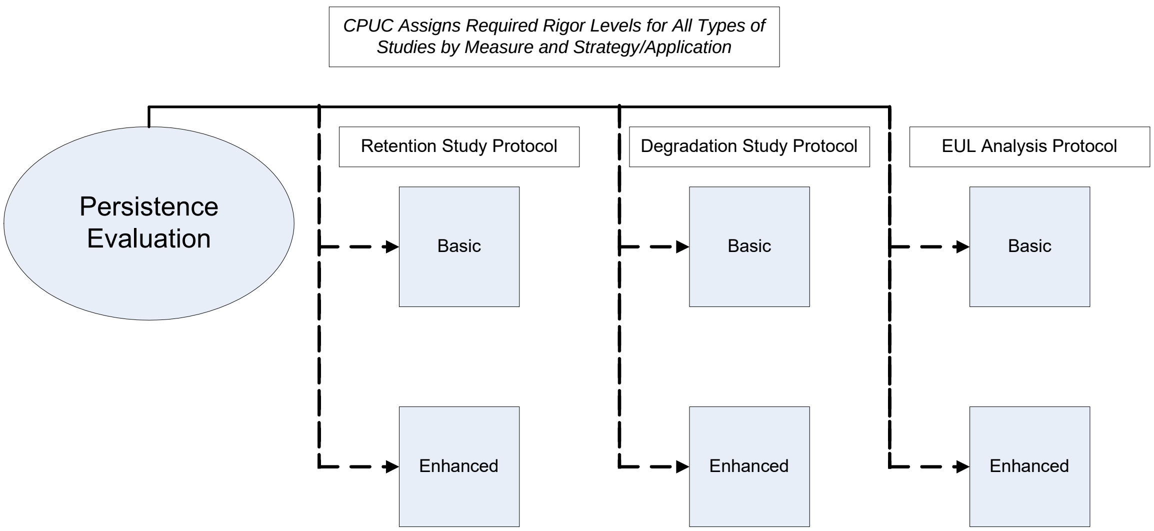

Figure 8. Protocols and Rigor Levels for EUL Evaluations

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 <u>within 30%</u> 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 <u>power analysis at a power of at least 0.7</u> 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 as 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	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.

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:

$$Pi = F(Z_i) = F(\alpha + \beta X_i) = 1/(1 + e^{-(\alpha + \beta X_i)})$$

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.

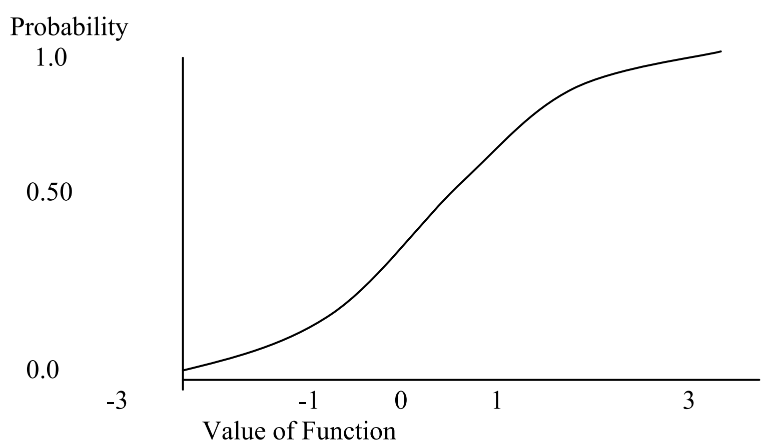

Figure 9. Cumulative Logistic Function

The logistic function that best fits the engineering observations described above relies upon a logistic function of time for identifying the EUL. This is:

$$F(Z_i) = 1\text{-}[1/(1 + e^{-(t + EUL)\ b})$$

With the survival function as in Figure 10.

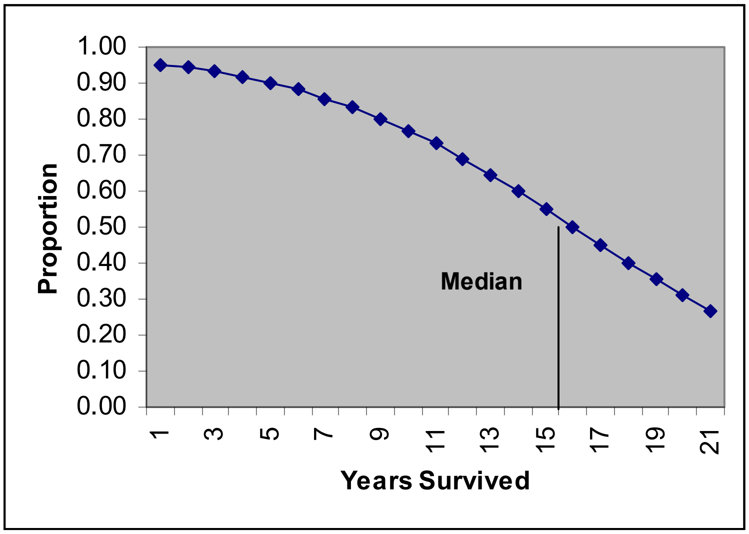

Figure 10. Logistic Survival Function with EUL=15 and b=0.2

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 tagging[87] 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.* 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 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 a 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 Prahl, 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.

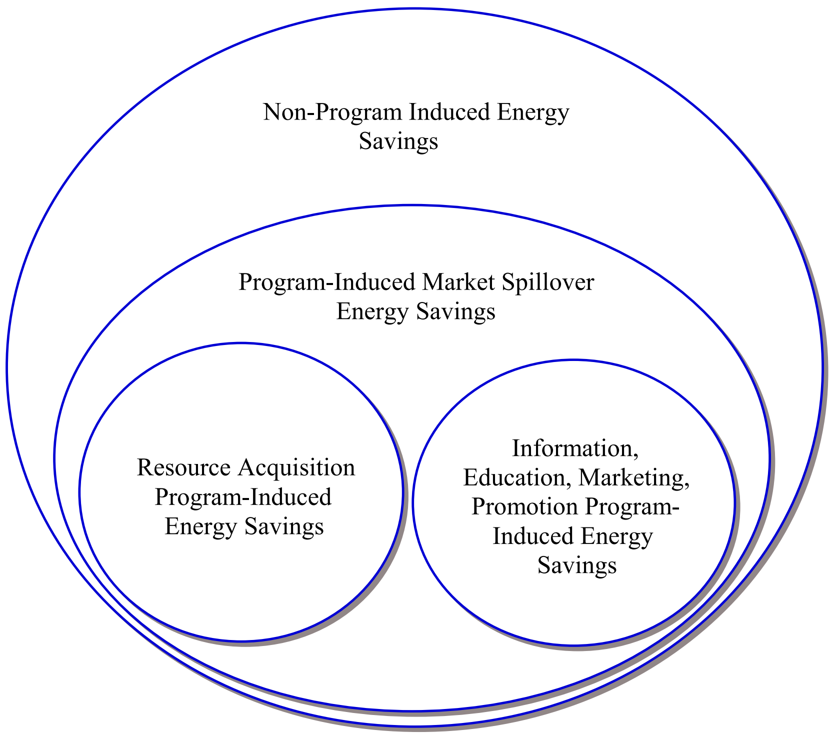

Figure 11. Sources of Energy Efficiency Changes in the Market

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 targets[117] 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 Impacts[120]

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 error[127]. 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 chose 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.

    - 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:
    - Engineering model bias – systematic under- or over-prediction of effects of a measure by an engineering model;
    - 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;
    - Deemed parameter bias – systematic deviation in a deemed parameter used in an engineering model;

- Meter bias – systematic error in meter and/or sensor;
- Sensor placement bias – systematic over- or under-prediction of measured quantity due to sensor placement (could be combined with above); and
- 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 indefinably 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
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	**June of 2006 Evaluation Projected Demand Impacts** (MW average weekday across periods)										
						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
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 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												
						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																	

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 **Program ID#: 1234-06** **All Program Measures** **Climate Zone: All Program Covered Climate Zones**		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
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 Program ID#: 1234-06 All Program Measures Climate Zone: All Program Covered Climate Zones		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
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 (use CPUC approved standard measure descriptions)	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
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 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

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
		Tons	High efficiency packaged AC and heat pumps, high efficiency chillers, waterside economizers, evaporative ventilation air pre-coolers
Lighting	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

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 Terms[141]

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-ft^2) or with cooling equipment that has a capacity exceeding 10 Btus/(hr-ft^2). 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×ft^2), 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×ft^2) 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 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.

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 (SHGG) - 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 though 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 Protocol[142] 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 Precision™ and is offered for sale by Biostat™ (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 Precision™ 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

	Power	
Effect	**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 Precision™*. 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	1. Participant and non-participant analysis of utility consumption data that addresses the issue of self-selection. 2. 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. 3. Econometric or discrete choice[149] 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 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 contactor 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 contactor 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.

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?).