Model Energy Efficiency Program Impact Evaluation Guide

A RESOURCE OF THE NATIONAL ACTION PLAN FOR ENERGY EFFICIENCY

NOVEMBER 2007

About This Document

This *Model Energy Efficiency Program Impact Evaluation Guide* is provided to assist gas and electric utilities, utility regulators, and others in the implementation of the recommendations of the National Action Plan for Energy Efficiency (Action Plan) and the pursuit of its longer-term goals.

This Guide describes a structure and several model approaches for calculating energy, demand, and emissions savings resulting from facility (non-transportation) energy efficiency programs that are implemented by cities, states, utilities, companies, and similar entities. By using best practices and consistent procedures, evaluations can support the adoption, continuation, and expansion of efficiency programs.

The primary audience for this Guide is energy efficiency program designers and evaluators looking for guidance on the evaluation process and key issues relating to documenting energy and demand savings, documenting avoided emissions, and comparing demand- and supply-side resources. Introductory portions and Appendix C are also intended for policy-makers seeking information about the basic principles of efficiency evaluation.

Model Energy Efficiency Program Impact Evaluation Guide

A RESOURCE OF THE NATIONAL ACTION PLAN FOR ENERGY EFFICIENCY

NOVEMBER 2007

The *Model Energy Efficiency Program Impact Evaluation Guide* is a product of the National Action Plan for Energy Efficiency Leadership Group and does not reflect the views, policies, or otherwise of the federal government. The role of the U.S. Department of Energy and U.S. Environmental Protection Agency is limited to facilitation of the Action Plan.

This document was final as of December 2007 and incorporates minor modifications to the original release.

If this document is referenced, it should be cited as:

National Action Plan for Energy Efficiency (2007). *Model Energy Efficiency Program Impact Evaluation Guide*. Prepared by Steven R. Schiller, Schiller Consulting, Inc. <www.epa.gov/eeactionplan>

For More Information

Regarding the *Model Energy Efficiency Program Impact Evaluation Guide*, please contact:

Niko Dietsch
U.S. Environmental Protection Agency
Office of Air and Radiation
Climate Protection Partnerships Division
Tel: (202) 343-9299
E-mail: dietsch.nikolaas@epa.gov

Regarding the National Action Plan for Energy Efficiency, please contact:

Stacy Angel
U.S. Environmental Protection Agency
Office of Air and Radiation
Climate Protection Partnerships Division
Tel: (202) 343-9606
E-mail: angel.stacy@epa.gov

Larry Mansueti
U.S. Department of Energy
Office of Electricity Delivery and Energy Reliability
Tel: (202) 586-2588
E-mail: lawrence.mansueti@hq.doe.gov

or visit www.epa.gov/eeactionplan

Table of Contents

Table of Contents (continued)

Table of Contents (continued)

List of Figures

List of Tables

List of Sidebars

List of Sidebars (continued)

List of Abbreviations and Acronyms

A

Action Plan	National Action Plan for Energy Efficiency
ASHRAE	American Society of Heating, Refrigerating, and Air Conditioning Engineers

B

BAU	business as usual
BM	build margin (for electricity generating units)

C

CDD	cooling degree day
CFL	compact fluorescent light bulb
CHP	combined heat and power

D

DEER	Database for Energy Efficiency Resources (California)
DOE	U.S. Department of Energy
DSM	demand-side management

E

ECM	energy conservation measure
EE	energy efficiency
EEM	energy efficiency measure
EM&V	evaluation, measurement, and verification
EPA	U.S. Environmental Protection Agency
ER	emission rate
EUL	effective useful life

F

FEMP	Federal Energy Management Program

G

GHG	greenhouse gas

H

HDD	heating degree day
HHV	high heating value
HVAC	heating, ventilation, and air conditioning

I

IPMVP	International Performance Measurement and Verification Protocol
ISO	independent system operator or International Organization for Standardization

K

kW	kilowatt
kWh	kilowatt-hour

L

lb	pound

M

M&V	measurement and verification
MW	megawatt
MWh	megawatt-hour

List of Abbreviations and Acronyms (continued)

N

NEB	non-energy benefits
NTGR	net-to-gross ratio

O

OM	operating margin (for electricity generating units)

Q

QAG	quality assurance guideline

T

TBE	theory-based evaluation
T&D	transmission and distribution

Acknowledgements

The Model Energy Efficiency Program Impact Evaluation Guide is a key product of the Year Two Work Plan for the National Action Plan for Energy Efficiency. This work plan was developed based on feedback received from the Action Plan Leadership Group members and observers during the fall of 2006. The work plan was further refined during the March 2007 Leadership Group meeting in Washington, D.C.

In addition to direction and comment by the Leadership Group, this Guide was prepared with highly valuable input from the Advisory and Technical Groups. Steven R. Schiller of Schiller Consulting, Inc., served as project manager and primary author to the Guide, under contract to the U.S. Environmental Protection Agency (EPA).

The Advisory Group members are:

- Commissioner Dian Grueneich, California Public Utilities Commission, co-chair
- Diane Munns, Edison Electric Institute, co-chair
- Chris James (formerly with the Connecticut Department of Environmental Protection)
- Rick Leuthauser, MidAmerican Energy Company
- Jan Schori, Sacramento Municipal Utility District
- Peter Smith (formerly with New York State Energy Research and Development Agency)

The Technical Group members are:

- Steve Schiller, Schiller Consulting (project manager and primary author)
- Derik Broekhoff, World Resources Institute
- Nick Hall, TecMarket Works
- M. Sami Khawaja, Quantec
- David Sumi, PA Consulting
- Laura Vimmerstedt, National Renewable Energy Laboratory
- Edward Vine, Lawrence Berkeley National Laboratory

In addition, the following professionals provided review and comments on drafts of the Guide: Joel Bluestein (EEA/ICF), Marian Brown (Southern California Edison), Cherie Gregoire (NYSERDA), Glenn Cannon (Waverly Light and Power), Ollie Frazier III (Duke Energy), Wilson Gonzalez (Ohio OCC), Leonard Haynes (Southern Company), Debra Jacobson (DJ Consulting), Ruth Kiselewich (Baltimore Gas & Electric Company), Sam Loudenslager (Arkansas PSC), Julie Michaels (Northeast Energy Efficiency Partnership), David Nemtzow (consultant), Monica Nevius (Consortium For Energy Efficiency), Anne Arquit Niederberger (Policy Solutions), Larry Pakenas (NYSERDA), Karen Penafiel (BOMA International), Alison Silverstein (consultant), Andrew Spahn (National Council on Electricity Policy), Rick Tempchin (Edison Electric Institute), Carol White (National Grid), and Vicki Wood (Sacramento Municipal Utility District).

The U.S. Department of Energy (DOE) and EPA facilitate the National Action Plan for Energy Efficiency, including this Guide. Key staff include Larry Mansueti (DOE Office of Electricity Delivery and Energy Reliability), Dan Beckley (DOE Office of Energy Efficiency and Renewable Energy), and Kathleen Hogan, Niko Dietsch, Stacy Angel, and Katrina Pielli (EPA Climate Protection Partnership Division).

Eastern Research Group, Inc., provided technical review, copy editing, graphics, and production services.

Executive Summary

This Model Energy Efficiency Program Impact Evaluation Guide provides guidance on model approaches for calculating energy, demand, and emissions savings resulting from energy efficiency programs. The Guide is provided to assist in the implementation of the National Action Plan for Energy Efficiency's five key policy recommendations for creating a sustainable, aggressive national commitment to energy efficiency.

Importance of Energy Efficiency Evaluation

Improving energy efficiency in our homes, businesses, schools, governments, and industries—which consume more than 70 percent of the natural gas and electricity used in the country—is one of the most constructive, cost-effective ways to address the challenges of high energy prices, energy security and independence, air pollution, and global climate change. Despite these benefits and the success of energy efficiency programs in some regions of the country, energy efficiency remains critically under utilized in the nation's energy portfolio. It is time to take advantage of more than two decades of experience with successful energy efficiency programs, broaden and expand these efforts, and capture the savings that energy efficiency offers. Program evaluation that is based on credible and transparent model methods is a key component of the solution.

Evaluation involves real time and/or retrospective assessments of the performance and implementation of a program. There are two key objectives of evaluations:

1. To document and measure the effects of a program and determine whether it met its goals with respect to being a reliable energy resource.
2. To help understand why those effects occurred and identify ways to improve current programs and select future programs.

Another objective can be to document compliance with regulatory requirements. Many energy efficiency evaluations are oriented toward developing retrospective estimates of energy savings attributable to a program, in a manner that is defensible in regulatory proceedings that are conducted to ensure that public funds are properly and effectively spent. However, the role of evaluation can go well beyond simply documenting savings to actually improving programs and providing a basis for future savings estimates. If applied concurrently with program implementation, evaluations can provide information in real time to allow for as-needed course corrections. In summary, evaluation fosters more effective programs and justifies increased levels of energy efficiency investment. Perhaps the imperative for conducting evaluation is best described by John Kenneth Galbraith and William Edwards Deming: "*Things that are measured tend to improve.*"

There are three different types of evaluations:

1. **Impact evaluations** determine the impacts (e.g., energy and demand savings) and co-benefits (e.g., avoided emissions, health benefits, job creation, energy security, transmission/distribution benefits, and water savings) that directly result from a program. Impact evaluations also support cost-effectiveness analyses aimed at identifying relative program costs and benefits.
2. **Process evaluations** assess program delivery, from design to implementation, in order to identify bottlenecks, efficiencies, what worked, what did not work,

constraints, and potential improvements. Timeliness in identifying opportunities for improvement is essential to making corrections along the way.

3. **Market effects evaluations** estimate a program's influence on encouraging future energy efficiency projects because of changes in the energy marketplace. These evaluations are primarily, but not exclusively, used for programs with market transformation elements and objectives.

The Role of This Guide

This Guide has been developed to assist parties in implementing the five key policy recommendations of the National Action Plan for Energy Efficiency. (See page 1-2 for a full list of options to consider under each Action Plan recommendation.) The Action Plan was released in July 2006 as a call to action to bring diverse stakeholders together at the national, regional, state, or utility level in order to foster the discussions, decision-making, and commitments necessary to take investment in energy efficiency to a new level.

This Guide supports the Action Plan recommendation to "make a strong, long-term commitment to implement cost-effective energy efficiency as a resource." A key option to consider under this recommendation is developing robust evaluation, measurement, and verification procedures. The model approaches described herein offer a set of options and an information resource for entities seeking to support the adoption, continuation, and expansion of energy efficiency programs.

The specific types of evaluations conducted are determined by the program goals and the objectives of those responsible for implementing and overseeing the programs. This Guide focuses on *impact evaluations* for programs designed to directly reduce energy consumption, demand, and air emissions. These programs are typically called *resource acquisition* programs, although other types of programs, such as market transformation programs, may also be assessed using impact evaluations. The efficiency programs considered here are those designed for facility or stationary (e.g., home, commercial building, factory) improvements, as opposed to transportation sector improvements.

The objective of this Guide is to provide a framework that jurisdictions and organizations can use to define their "institution-specific" or "program/portfolio-specific" evaluation requirements. To this end, the Guide defines a standard evaluation planning and implementation process, describes several standard approaches that can be used for calculating savings, defines terms, provides advice on key evaluation issues, and lists efficiency evaluation resources. While each jurisdiction, or entity, will need to define its own policy requirements, this Guide provides a structure for applying consistent approaches and definitions. This can facilitate the implementation of "cross-border" programs to establish energy efficiency as a priority resource or as a greenhouse gas mitigation option.

The audience for this Guide is energy efficiency program designers and evaluators looking for guidance, resources, and references on the evaluation process and key issues relating to (a) documenting energy and demand savings and (b) documenting avoided emissions. Introductory portions of this Guide are also intended for policy-makers seeking information about the basic principles of impact evaluation. Readers looking only for basics may want to read only this executive summary and the first few chapters, and perhaps refer to the appendices for overviews of other evaluation types, definitions, and references. Some readers who are new to evaluation assignments may read the entire document, while others may benefit from focusing on the evaluation planning chapter (Chapter 7) and using the rest of the document as a reference.

Overview of the Program Impact Evaluation Process

The basic steps in the impact evaluation process are:

- Setting the evaluation objectives in the context of the program policy objectives.
- Selecting an evaluation approach and preparing a program evaluation plan that takes into account the critical evaluation issues.
- Implementing the evaluation and determining program impacts, such as energy and demand savings and avoided emissions.
- Reporting the evaluation results and, as appropriate, working with program administrators to implement recommendations for current or future program improvements.

This Guide is about program, versus project, evaluation. In this context, a project is a single activity at one location (for example, an energy-efficient lighting retrofit in an office building). A program is a group of projects with similar characteristics that are installed in similar applications, such as a utility program to install energy-efficient lighting in commercial buildings, a company's program to install energy management system in all of its stores, or a state program to improve the efficiency of its public buildings. Programs are typically evaluated using a sample (versus a census) of projects, with the results systematically applied to the entire program "population" of projects. Sampling is one of the issues discussed in the Guide.

The three impact evaluation results that are typically reported are:

- **Estimates of gross savings.** Gross energy (or demand) savings are the change in energy consumption or demand that results directly from program-promoted actions (e.g., installing energy-efficient lighting) taken by program participants regardless of the extent or nature of program influence on their actions.
- **Estimates of net savings.** Net energy or demand savings refer to the portion of gross savings that is attributable to the program. This involves separating out the impacts that are a result of other influences, such as consumer self-motivation. Given the range of influences on consumers' energy consumption, attributing changes to one cause (i.e., a particular program) or another can be quite complex.
- **Estimates of co-benefits.** A co-benefit commonly documented and reported is avoided air emissions— the air pollution or greenhouse gases that would have been emitted if more energy had been consumed in the absence of the energy efficiency program. These emissions can be from combustion of fuels at an electrical power plant or from combustion of heating fuels, such as natural gas and fuel oil, at a project site. Other co-benefits can be positive or negative; examples are comfort and productivity improvements, job creation, and increased maintenance costs due to unfamiliarity with new energy-efficient equipment.

It is important to note that energy and demand savings, and avoided emissions, cannot be directly measured. Instead, savings are determined by comparing energy use and demand after a program is implemented (the reporting period) with what would have occurred had the program not been implemented (the baseline). The baseline and reporting period energy use and demand are compared using a common set of conditions (e.g., weather, operating hours, building occupancy). These are then adjusted so that only program effects are considered when determining savings. Avoided emissions and other co-benefits can then be calculated using the energy savings values and other relevant information.

Note that each of the above bullets defines an "estimate." This is because the nature of efficiency evaluation involves measuring energy consumption. The difference between (a) actual energy consumption and (b) what energy consumption would have occurred during the same period had the efficiency measures not been installed, is an *estimate* of energy (and demand) savings. The energy that would have been consumed during that same time *was not*, and so must be estimated rather than measured.

As indicated, a key objective of program evaluation is to produce an estimate of energy and demand savings (and, as desired, associated co-benefits). However, the value of the estimates as a basis for decision-making can be called into question if their sources and level of accuracy are not analyzed and described. Therefore, evaluation results, like any estimate, should be reported as "expected values" with an associated level of uncertainty. Minimizing uncertainty and balancing evaluation costs with the value of the evaluation information are at the heart of the evaluation process.

Implementing the impact evaluation process for determining energy and demand savings, and avoided emissions, involves:

1. Determining gross program savings using one of the following approaches:

   a. One or more measurement and verification (M&V) methods, from the IPMVP,[1] are used to determine the savings from a sample of projects. These savings are then applied to all of the projects in the program.

   b. Deemed savings, based on historical and verified data, are applied to conventional energy efficiency measures implemented in the program.

   c. Statistical analyses of large volumes of metered energy usage data are conducted.

   In some cases these approaches are combined, particularly the deemed savings and M&V approaches.

2. Converting gross program savings to net energy savings using a range of possible considerations. The primary, but not exclusive, considerations that account for the difference between net and gross savings are free riders (i.e., those who would have implemented the same or similar efficiency projects without the program now or in the near future) and participant and non-participant spillover. Non-participant spillover is defined as savings from efficiency projects implemented by those who did not directly participate in a program, but which nonetheless occurred due to the influence of the program. Participant spillover is defined as additional energy efficiency actions taken by program participants as a result of program influence, but actions that go beyond those directly subsidized or required by the program. Net savings are determined using one of the following approaches:

   a. Self-reporting surveys in which information is reported by participants and non-participants without independent verification or review.

   b. Enhanced self-reporting surveys in which self-reporting surveys are combined with interviews and documentation review and analysis.

   c. Statistical models that compare participants' and non-participants' energy and demand patterns, their knowledge about efficiency options, and/or the trade-offs they are willing to make between efficiency options and the costs of purchasing and installing them.

   d. Stipulated net-to-gross ratios (ratios that are multiplied by the gross savings to obtain an estimate of net savings) that are based on historic studies of similar programs.

3. Calculating avoided emissions by either (a) applying emission factors (e.g., pounds of CO_2 per MWh) to net energy savings or (b) using emissions scenario analyses (e.g., using computer models to estimate the difference in emissions from grid-connected power plants with and without the reduced electricity consumption associated with an efficiency program). Within these two categories, a variety of approaches can be used to calculate emission factors or prepare scenarios analyses ranging from using a simple annual average emission factor to prepare detailed hourly calculations of displaced energy

sources and their emissions. However, the question of whether emissions are actually avoided depends on whether the energy savings are truly additional to what would have occurred without the program's influences, whether all significant emissions sources associated with a program were taken into account, and the scheme under which any affected emission sources may be regulated.

Evaluation Characteristics and Evaluation Planning

The following practices are commonly adopted as part of the evaluation process:

- The evaluation process is integral to a typical cyclic planning-implementation-evaluation process. Therefore, evaluation planning is part of the program planning process so that the evaluation effort can support program implementation, including the alignment of implementation and evaluation budgets and schedules, and can provide evaluation results in a timely manner to support existing and future programs.
- Evaluation budgets and resources are adequate to support, over the entire evaluation time period, the evaluation goals and the level of quality (certainty) expected in the evaluation results.
- Evaluations use the planning and implementation structure described in this Guide, as well as the definitions provided for evaluation terms.
- Energy and demand savings calculations follow one or more of the approaches defined in this Guide for net and gross savings.
- Evaluations are complete, transparent, relevant, consistent, and balanced in risk management between certainty of results and costs to achieve the results. They also follow the guiding principles defined by the American Evaluation Association, which are listed in this Guide (see Section 3.8).

With the above characteristics in mind, individual entities can define their own policy-specific program evaluation requirements. These requirements are determined by the program objectives, regulatory mandates (if any), expectations for quality of the evaluation results, intended uses of the evaluation results, and other factors that can vary across jurisdictions and programs. In this Guide, seven key evaluation planning issues are defined and discussed to help define policy-specific program evaluation requirements. These are:

1. Defining evaluation goals and scale, including deciding which program benefits to evaluate.
2. Setting a time frame for evaluation and reporting expectations.
3. Setting a spatial boundary[2] for evaluation (i.e., what energy uses, emission sources, etc., the analyses will include).
4. Defining a program baseline, baseline adjustments, and data collection requirements.
5. Establishing a budget in the context of expectations for the quality of reported results.
6. Selecting impact evaluation approaches for calculating gross and net savings and avoided emissions.
7. Selecting the individual or organization that will conduct the evaluation.

The issues above are listed in what can be considered a sequential process, however many are interrelated and the overall planning process is iterative. After each of these issues is addressed individually, the results can be compiled into a formal evaluation plan.

In conclusion, this Guide can be used at the onset of program planning to initiate a parallel evaluation planning effort. Doing so will help evaluators take an integral role in the program's success and help those who are implementing the program understand the parameters under which they will be evaluated and what information they are expected to provide, and receive from, the evaluation.

Notes

1. Measurement and verification is the process of using measurements to reliably determine actual savings created within an individual facility. IPMVP is the International Performance Measurement and Verification Protocol (available at <http://www.evo-world.org>). The IPMVP is a measurement and verification protocol for projects, whereas this Guide focuses on programs, which are collections of similar projects.

2. Spatial boundary refers to "how big a circle is going to be drawn around" the energy efficiency measures being evaluated. Is the analysis only going to be on the affected equipment, the whole facility, or perhaps even the entire generation, transmission, and distribution system?

1: Introduction

Improving the energy efficiency of homes, businesses, schools, governments, and industries—which together consume more than 70 percent of the natural gas and electricity used in the United States—is one of the most cost-effective ways to address the challenges of high energy prices, energy security and independence, air pollution, and global climate change. Mining this efficiency could help us meet on the order of 50 percent or more of the expected growth in U.S. consumption of electricity and natural gas in the coming decades, yielding many billions of dollars in saved energy bills and avoiding significant emissions of greenhouse gases and other air pollutants.[1]

Recognizing this large opportunity, more than 60 leading organizations representing diverse stakeholders from across the country joined together to develop the National Action Plan for Energy Efficiency. The Action Plan identifies many of the key barriers contributing to underinvestment in energy efficiency; outlines five key policy recommendations for achieving all cost-effective energy efficiency, focusing largely on state-level energy efficiency policies and programs; and provides a number of options to consider in pursuing these recommendations (Figure 1-1). As of November 2007, nearly 120 organizations have endorsed the Action Plan recommendations and made public commitments to implement them in their areas. Effective energy efficiency program evaluation is key to making the Action Plan a reality.

1.1 About the Guide

Guide Objective

After reading this Guide, the reader will be able to define the basic objectives, structure, and evaluation approaches that can be used to conduct program-specific impact evaluation. Depending on experience level, the reader may be able to prepare a complete program impact evaluation plan. Appendix E provides a list of references that can also assist with this process.

This Guide describes a structure and several model approaches for calculating energy, demand, and emissions savings from energy efficiency programs. By adhering to best practices and standard procedures, stakeholders can use program evaluation as an effective tool to support the adoption, continuation, and expansion of energy efficiency programs.

The Action Plan's Leadership Group (see Appendix A for a list of group members) identified energy efficiency program evaluation, measurement, and verification (EM&V) as an area where additional guidance is needed to help parties pursue the recommendations and meet their commitments to energy efficiency. Specifically, this Guide supports the Action Plan recommendation to "Make a strong, long-term commitment to implement cost-effective energy efficiency as a resource." A key option to consider under this recommendation is to develop robust measurement and verification procedures that support the adoption, continuation, and expansion of energy efficiency programs.

Further, two recent surveys of the energy efficiency industry indicated a need for guidance documents that foster best practices for evaluation and promote consistent evaluations of energy efficiency programs (NEEP, 2006; Schiller Consulting, 2007). This Guide fills the identified gaps by providing:

- A model impact evaluation process that individual jurisdictions (e.g., states, utilities) can use to establish their own evaluation requirements.
- Policy-neutral[2] descriptions and guidance for conducting impact evaluations of resource acquisition programs.

Figure 1-1. National Action Plan for Energy Efficiency Recommendations and Options

Recognize energy efficiency as a high-priority energy resource.

Options to consider:

- Establishing policies to establish energy efficiency as a priority resource.
- Integrating energy efficiency into utility, state, and regional resource planning activities.
- Quantifying and establishing the value of energy efficiency, considering energy savings, capacity savings, and environmental benefits, as appropriate.

Make a strong, long-term commitment to implement cost-effective energy efficiency as a resource.

Options to consider:

- Establishing appropriate cost-effectiveness tests for a portfolio of programs to reflect the long-term benefits of energy efficiency.
- Establishing the potential for long-term, cost-effective energy efficiency savings by customer class through proven programs, innovative initiatives, and cutting-edge technologies.
- Establishing funding requirements for delivering long-term, cost-effective energy efficiency.
- Developing long-term energy saving goals as part of energy planning processes.
- Developing robust measurement and verification procedures.
- Designating which organization(s) is responsible for administering the energy efficiency programs.
- Providing for frequent updates to energy resource plans to accommodate new information and technology.

Broadly communicate the benefits of and opportunities for energy efficiency.

Options to consider:

- Establishing and educating stakeholders on the business case for energy efficiency at the state, utility, and other appropriate level, addressing relevant customer, utility, and societal perspectives.
- Communicating the role of energy efficiency in lowering customer energy bills and system costs and risks over time.
- Communicating the role of building codes, appliance standards, and tax and other incentives.

Provide sufficient, timely, and stable program funding to deliver energy efficiency where cost-effective.

Options to consider:

- Deciding on and committing to a consistent way for program administrators to recover energy efficiency costs in a timely manner.
- Establishing funding mechanisms for energy efficiency from among the available options, such as revenue requirement or resource procurement funding, system benefits charges, rate-basing, shared-savings, and incentive mechanisms.
- Establishing funding for multi-year periods.

Modify policies to align utility incentives with the delivery of cost-effective energy efficiency and modify ratemaking practices to promote energy efficiency investments.

Options to consider:

- Addressing the typical utility throughput incentive and removing other regulatory and management disincentives to energy efficiency.
- Providing utility incentives for the successful management of energy efficiency programs.
- Including the impact on adoption of energy efficiency as one of the goals of retail rate design, recognizing that it must be balanced with other objectives.
- Eliminating rate designs that discourage energy efficiency by not increasing costs as customers consume more electricity or natural gas.
- Adopting rate designs that encourage energy efficiency by considering the unique characteristics of each customer class and including partnering tariffs with other mechanisms that encourage energy efficiency, such as benefit-sharing programs and on-bill financing.

Source: National Action Plan for Energy Efficiency, 2006.

- A list of other reference documents and resources on energy efficiency evaluation.
- Information on calculating avoided emissions from energy efficiency programs.

Jurisdictions and organizations can use this Guide as both a primer on efficiency impact evaluation and a framework to define their own institution-specific, program-specific, or portfolio-specific evaluation requirements. While each jurisdiction or entity will need to define its own policy requirements, this Guide provides a structure, evaluation approaches, and definitions that can be applied to a variety of policy requirements. If applied consistently, the approaches described in this Guide could ease the implementation of "cross-border" programs to establish energy efficiency as a priority resource or as a greenhouse gas mitigation option.

1.2 Subjects Covered in This Guide

This Guide focuses on evaluating the impact—i.e., the energy, demand, and emissions savings—of energy efficiency programs implemented in facilities.[3] Therefore, the Guide helps determine the fuel oil, natural gas, and electricity savings from programs that encourage lighting, space conditioning, process approaches, and similar energy efficiency strategies in residential, commercial, and industrial facilities. Also addressed are the avoided emissions associated with these energy savings.

The Guide is intended to assist in the evaluation of programs for which energy and demand savings are the primary objectives (commonly referred to as "resource acquisition" programs), although other types of programs may be assessed using impact evaluations. Appendix C briefly discusses evaluation approaches for market transformation, codes and standards, and education programs. It also describes process, market, and cost-effectiveness evaluations.

This Guide lays out a basic evaluation structure, highlighting issues that need to be addressed in order to prepare a jurisdiction-specific evaluation plan or protocol for a single program or portfolio of programs.[4] These issues include:

1. Defining evaluation goals and scale. (This includes deciding which program benefits to evaluate.)
2. Setting a time frame for evaluation and reporting expectations.
3. Setting a spatial boundary for evaluation.
4. Defining a program baseline, baseline adjustments, and data collection requirements.
5. Establishing a budget in the context of expectations for the quality of reported results.
6. Selecting impact evaluation approaches for gross and net savings calculations, and avoided emissions calculations.
7. Selecting who (or which type of organization) will conduct the evaluation.

Planning Issues

While reading this Guide's first six chapters, the reader should keep in mind the seven "evaluation planning" issues listed in Section 1.2. Chapter 7 addresses these issues in more detail and describes how material from previous chapters can be used to prepare an evaluation plan.

It is also important to indicate what the Guide does *not* cover:

- The Guide is not sufficiently detailed to be the only resource for planning or conducting evaluations of specific programs. Rather, it provides high-level guidance, identifies issues, and direct users to resources for defining policy- and program-specific requirements and details. For example, it does not describe specific data collection and analysis options, although Appendix E does list documents where this information can be found for various program types and technologies.

- The Guide is not intended for use in assessing the savings and benefits from a *future* energy efficiency program, but rather to inform on what has been, is being, or is projected to be accomplished with an existing program.

1.3 How to Use This Guide

In practical terms, evaluation planners can use this Guide to:

- Define the questions and hypotheses that the evaluation effort is intended to answer.
- Identify appropriate evaluation approaches and methods that minimize uncertainty while meeting budget constraints.
- Set realistic expectations among the evaluation process stakeholders regarding the nature and practical value of results to be delivered, as well as the expected quality of quantitative estimates of program impacts.
- Set appropriate schedules and budgets that reflect the level of certainty expected in the results.

In addition, introductory portions of this Guide are also intended for policy-makers seeking information about the basic principles of impact evaluation.

The intended audience is:

- Program and evaluation managers looking for basic guidance—or a "roadmap"—on process and key issues relating to:
  - Documenting energy and demand savings.
  - Documenting avoided emissions.
  - Comparing demand- and supply-side resources.
- Program designers looking to understand how their programs will be evaluated.
- Policy-makers and regulators looking for a basic understanding of evaluation objectives, processes, and issues.
- Members of the energy efficiency community looking for:
  - Common terminology definitions.
  - A central reference that provides guidance, but also lists publicly available best practices resources.
  - An understanding of the mechanisms for determining the potential value of energy efficiency as an emissions avoidance strategy.

Using This Guide

Policy-makers and those looking for the "basics": Read the Executive Summary and first few chapters, and perhaps refer to the appendices for overviews of other evaluation types, definitions, and references.

Experienced evaluation planners: Go straight to the planning chapter (Chapter 7) and use the rest of the document as a reference.

Readers new to evaluation or energy efficiency: Read the entire document.

Table 1-1 to the right also summarizes the contents and intended readers for each part of the Guide.

1.4 Source Documents

The information in this document is a summary of definitions, approaches, and best practices developed over the last 30 years of energy efficiency program implementation and evaluation. This experience and expertise is documented in numerous guides, protocols, papers, and reports. The key documents that were used in the development of the Guide are:

- 2007 International Performance Measurement and Verification Protocol (IPMVP).
- 2006 California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals.
- 2000 FEMP M&V Guidelines.

Table 1-1. Model Energy Efficiency Program Impact Evaluation Guide Overview

Document Element	Titles	Contents and Intended Audience
Part 1	*Executive Summary*	Summarizes importance and types of evaluations, the impact evaluation process, key issues, and evaluation planning. **Intended for all readers.**
Part 2	*Chapter 1: Introduction* *Chapter 2: Energy Efficiency Program Evaluation* *Chapter 3: Impact Evaluation Basics*	Provides basics of energy efficiency evaluation. **Chapters 2 and 3 are intended for readers who want an overview of evaluation and the key aspects of impact evaluation.**
Part 3	*Chapter 4: Calculating Gross Energy and Demand Savings* *Chapter 5: Calculating Net Energy and Demand Savings* *Chapter 6: Calculating Avoided Air Emissions*	Provides details on the process and approaches for quantifying energy and demand savings, and avoided emissions, from energy efficiency programs. **Intended for readers whose programs are to be evaluated, evaluators, and managers/regulators of evaluation activities.**
Part 4	*Chapter 7: Planning an Impact Evaluation*	This chapter "brings it all together" and describes how the information described in earlier Chapters can be utilized to plan an evaluation effort. **Also intended for readers whose programs are to be evaluated, evaluators, and managers/regulators of evaluations.** ***Some readers with background in evaluation may want to go directly to this chapter.***
Part 5	*Appendix A: National Action Plan for Energy Efficiency Leadership Group* *Appendix B: Glossary* *Appendix C: Other Evaluation Types* *Appendix D: Uncertainty* *Appendix E: Resources* *Appendix F: Renewables and Combined Heat and Power Project Measurement and Verification*	These Appendices provide resources and further background on evaluation issues. **Intended for readers interested in specialty subjects or reference materials.** **Appendix B, the glossary, and Appendix C may be of interest to policy-makers.** **Appendix C summarizes the various types of efficiency programs and the ways these programs can be evaluated.**

- 2004 California Public Utilities Commission (CPUC) Evaluation Framework.
- 2002 ASHRAE Guideline 14 Measurement of Energy and Demand Savings.

More information on these documents and other evaluation resources is contained in Appendix E.

1.5 Structure of the Guide

This Guide primarily covers *impact evaluations* (i.e., determining the energy, demand, and emissions savings that directly result from a program) and is organized into five parts:

- The Executive Summary, which briefly describes the evaluation process outlined in this Guide.
- Chapters 1 through 3, which introduce this Guide, key energy efficiency concepts, and program impact evaluation concepts and basics.
- Chapters 4 through 6, the core of the Guide, which describe approaches for determining gross and net energy/demand savings, and avoided emissions, from energy efficiency programs.
- Chapter 7, which discusses the evaluation planning process and key evaluation planning issues as well as presenting some evaluation plan outlines that entities can use to prepare their own evaluation requirements.
- Appendices on terminology, references and resources, other types of program evaluations (process and market), evaluation statistics, and evaluation of combined heat and power (CHP) and renewable energy programs.

1.6 Development of the Guide

This Guide is a product of the Year Two Work Plan for the National Action Plan for Energy Efficiency. The Action Plan's Leadership Group formed an Advisory Group and a Technical Group to help develop the Guide. Steven R. Schiller of Schiller Consulting, Inc., was contracted to serve as project manager and primary author. Commissioner Dian Grueneich (California Public Utilities Commission) and Dian Munns (Executive Director of Retail Energy Services, Edison Electric Institute) co-chaired the Guide's Advisory Group.

Additional Advisory Group members include:

- Chris James (formerly with the Connecticut Department of Environmental Protection).
- Rick Leuthauser, MidAmerican Energy Company.
- Jan Schori, Sacramento Municipal Utility District.
- Peter Smith (formerly with New York State Energy Research and Development Agency).

The Technical Group members are:

- Steve Schiller, Schiller Consulting: project manager and primary author.
- Derik Broekhoff, World Resources Institute.
- Nick Hall, TecMarket Works.
- M. Sami Khawaja, Quantec: Appendix D author.
- David Sumi, PA Consulting.
- Laura Vimmerstedt, National Renewable Energy Laboratory.
- Edward Vine, Lawrence Berkeley National Laboratory.

1.7 Notes

1. See the *National Action Plan for Energy Efficiency* (2006), available at <http://www.epa.gov/cleanenergy/actionplan/report.htm>.
2. The Guide is "policy neutral" in that it can be applied to energy efficiency and emission avoidance programs irrespective of the programs' policy objectives or constraints.
3. The Guide does not cover transportation-related efficiency programs.
4. Since the Guide is a policy-neutral document, following it will not necessarily ensure that a program evaluation plan will be in compliance with regulatory or similar mandates. The entity-specific program plan must address any jurisdictional policy requirements.

2: Energy Efficiency Program Evaluation

Chapter 2 provides a brief overview of the importance of energy efficiency evaluation and describes the context in which it is conducted. The chapter also makes the distinction between evaluations for individual energy efficiency projects and multifaceted efficiency programs. Because this Guide focuses on program evaluation, additional background on program categories and related evaluation approaches is provided.

2.1 Importance of Evaluation

Evaluation is the process of determining and documenting the results, benefits, and lessons learned from an energy efficiency program. Evaluation results can be used in planning future programs and determining the value and potential of a portfolio of energy efficiency programs in an integrated resource planning process. It can also be used in retrospectively determining the performance (and resulting payments, incentives, or penalties) of contractors and administrators responsible for implementing efficiency programs.

Evaluation has two key objectives:

1. To document and measure the effects of a program and determine whether it met its goals with respect to being a reliable energy resource.
2. To help understand why those effects occurred and identify ways to improve current programs and select future programs.

Energy efficiency evaluations are conducted to estimate retrospective or real-time energy savings (versus predicted estimates) attributable to a program in a manner that is defensible in regulatory proceedings. However, evaluation should be viewed as one part of a continuous, and usually cyclic, process of program planning, implementation, and evaluation. Thus, the results of impact evaluation studies do not stand alone, but are used as inputs into planning and improving future programs.[1] Furthermore, rigorous evaluations help ensure that programs are cost-effective and savings are sustained over time.

There are several technical and policy barriers to the full use of cost-effective energy efficiency, and to the incorporation of efficiency programs into energy resource portfolios. One of these barriers is proving that energy efficiency "can be counted on." Consistent, complete, accurate, and transparent evaluation mechanisms for documenting energy savings and avoided emissions address this barrier. Indeed, having effective evaluation policies, processes, and trained personnel in place to document the energy and environmental benefits of energy efficiency programs is critical to the success of energy efficiency and climate mitigation programs that must prove their value and worthiness for continued investment.

Some Applications of Energy Efficiency Evaluation

- Utility-administered energy efficiency programs.
- Government efficiency programs, either for their own facilities or for private-sector incentive programs.
- Independent system operator (ISO) programs to reduce demand (e.g., a forward capacity market).
- Air-pollution and greenhouse gas mitigation programs that utilize efficiency.
- Private company programs.
- Energy service company contracts.

Why Conduct Evaluations?

The reasons to do an evaluation can be summarized in two words: improvement and accountability. Evaluations provide information that can help improve programs and demonstrate internal and external accountability for the use of resources.

Program evaluations provide timely information to improve program implementation, as well as the design of future programs and individual energy efficiency projects. They can answer the following questions:

- Are the program and the projects that make up the program achieving their goals? If so, how and why?
- How well has the program/project worked?
- What changes are needed to improve the program/project?
- What is the program's impact on actual projects and future projects?
- Should the program/project be replicated, adjusted, or cancelled?

An evaluation also indicates whether the "resource" can be relied upon. Knowing whether the efficiency program will reliably generate savings (e.g., MWh) is critical to the ability of existing and future programs to serve as an important part of an energy resource portfolio.

An evaluation also provides an understanding of:

- Program approaches that are most and least effective, and how to improve future programs.
- Where to focus for greater savings.
- Actual values that can be used in future estimates of benefits (e.g., estimates of energy savings per square foot of office space).

2.2 Defining Program Versus Project Evaluation

A program is a group of projects with similar technology characteristics that are installed in similar applications, such as a utility program to install energy-efficient lighting in commercial buildings, a company's program to install energy management system in all of its stores, or a state program to improve the efficiency of its public buildings. A portfolio is either (a) a collection of similar programs addressing the same market (e.g., a portfolio of residential programs), technology (e.g., motor efficiency programs), or mechanisms (e.g., loan programs) or (b) the set of all programs conducted by a particular entity, which could include programs that cover multiple markets, technologies, etc. This Guide covers program evaluation, though the basic concepts can be applied to a portfolio if the impacts of interactions between programs and savings estimates are considered. In this context, a project is a single activity at one location, such as an energy-efficient lighting retrofit in an office building. Programs are often evaluated using a sample (versus a census) of projects, with the results applied to the entire program "population" of projects.

2.3 Efficiency Program Categories

Energy efficiency programs are planned and coordinated actions designed for a specific purpose. These actions are usually made up of projects carried out at individual facilities, for example as part of a utility efficiency incentive program. There are many types of energy efficiency programs but no standard way of differentiating them—this Guide differentiates programs by their primary objectives:

- **Resource acquisition**—primary objective is to *directly* achieve energy and/or demand savings, and possibly avoid emissions, through specific actions.
- **Market transformation**—primary objective is to change the way in which energy efficiency markets operate (how manufacturers, distributors, retailers, consumers, and others sell and buy energy-related products and services), which tends to result in

energy and demand savings in a more *indirect* manner. To a large extent, all programs can be considered market transformation in that they involve changing how energy efficiency activities take place in the marketplace.

- **Codes and standards**—primary objective is to define and enforce mandated levels of efficiency in buildings and products.
- **Education and training**—primary objective is to inform consumers and providers about energy efficiency and encourage them to act on that information.
- **Multiple objective**—objectives can include some or all of the above listed objectives.

This Guide focuses on documenting the impacts of resource acquisition programs, including directly achieved energy and demand savings, and related emissions reductions. Appendix C briefly discusses evaluation of the other program categories listed above. It should be noted that while a program may have one primary objective there are often secondary objectives that are integral to program's overall success. This is frequently the case when resource acquisition and market transformation objectives are involved. With respect to evaluation, it is more important to focus on the performance goals to be assessed than on categorizing individual program types.

Energy efficiency is part of the general category of activities known as demand-side management (DSM). DSM programs are designed to encourage consumers to modify their level and pattern of energy usage. Another category of DSM is demand response (DR), defined by DOE as "changes in electric usage by end-use customers from their normal patterns in response to changes in the price of electricity over time, or to incentive payments designed to induce lower electricity use at times of high wholesale market prices or when system reliability is jeopardized" (DOE, 2006). DR programs employ rate design, customer incentives, and technology to enable customers to change their demand in response to system conditions or prices. Effective DR programs can improve system reliability and reduce capital costs associated with transmission and generation capacity investment by lowering overall peak demand. The Action Plan recognizes the value of using less energy at any time, including time of peak demand, through DR and peak shaving efforts. However, DR

NYSERDA Portfolio Evaluation Approach

The evaluation resources available for NYSERDA's New York Energy \$mart^{SM} Program are more limited than is common across most programs. In the traditional approach, single programs are evaluated, using any or several of the primary types of evaluation—impact, process, market, etc.—by either a single contracted evaluator, a single evaluator using a team of subcontractors, or a consulting firm. This can be effective when funds are sufficient, programs are evaluated one at a time, and those programs are essentially independent from one another.

In NYSERDA's case, there was concern that the traditional approach might be less useful given that its many programs are intended to work in tandem to meet the needs of multiple customers.

NYSERDA was also concerned that the traditional approach would not be sufficient, given available resources, to determine whether public policy goals set for the New York Energy $mart Program were being met.

To meet its unique needs, NYSERDA selected an evaluation approach that departs from the traditional method of focusing on a single program. NYSERDA hires teams of contractors that specialize in one type of evaluation, and then each team analyzes a suite of programs. At the end of an evaluation cycle, NYSERDA combines and integrates the results from each of the program evaluations and "rolls them up" to the portfolio level to provide an estimate of the overall effects of the portfolio (i.e., the whole of New York Energy $mart) and its progress toward achieving the public policy goals.

Provided by NYSERDA.

programs (a) may have relatively short-term effects on energy consumption, (b) may shift use from a time of high energy costs to a lower-cost time, but not reduce overall electricity use, and (c) may reduce energy use at high-cost times by paying for a reduction in the level of service provided.

Energy efficiency evaluation has a fairly long history, while DR evaluation is relatively new and appropriate methodologies are still under development.[2] While this Guide does not specifically address DR programs, the basic evaluation approaches and planning process explained here can be applied to DR with the understanding that the emphasis for DR program evaluation is demand savings. Demand savings definitions and evaluation techniques are highlighted in Section 3.2. Chapter 7 includes a sidebar on the ISO-New England demand resources program measurement and verification Guide, and Appendix E includes additional DR references.

2.4 Program Evaluation Categories

Evaluation involves retrospectively assessing the performance and implementation of a program. The following bullets describe three basic types of evaluations, all considered “ex post” because they analyze what has already occurred. The Guide focuses primarily on impact evaluations that quantify direct energy and capacity saving benefits. The other two evaluation types are summarized in more detail in Appendix C.

1. **Impact evaluations** determine the impacts (usually energy and demand savings) and co-benefits (such as avoided emissions health benefits, job creation, and water savings) that directly result from a program. All categories of energy efficiency programs can be assessed using impact evaluations, but they are most closely associated with resource acquisition programs.

2. **Process evaluations** assess how efficiently a program was or is being implemented, with respect to its stated objectives and potential for future improvement. All energy efficiency program categories can be assessed using process evaluations.

Program Planning and Evaluation

Evaluation is a retrospective process for determining how a program performed over a specific period of time (e.g., month, season, year). The Latin term *ex post* (meaning after the fact) is used to describe the typical evaluation process. This is in contrast to *a priori* (before the activity—postulated or prospective) analyses. Note though, that evaluations that produce results while the program is operating can be very useful. When possible, evaluations should be done within a program cycle so that feedback is frequent and systematic and benefits the existing program and informs the design of future programs and their evaluation.

For planning a future program, historical evaluation results can help with program design. However, for estimating how a program will perform, *potential studies* and *feasibility studies* are the typical analyses performed. Both of these studies look at what levels of savings are possible from technical, economic and market-acceptance perspectives. Potential studies are typically conducted on a market sector basis (e.g., residential, commercial, industrial sectors) and feasibility studies tend to be focused on specific customers that may be involved in a particular program. For more information on program planning, see National Action Plan for Energy Efficiency, 2007a. For more information on energy efficiency potential studies, see National Action Plan for Energy Efficiency, 2007b.

3. **Market effects evaluations** estimate a program’s influence on encouraging future energy efficiency projects because of changes in the marketplace. While all categories of programs can be assessed using market effects evaluations, they are primarily associated with market transformation programs that indirectly achieve impacts and resource acquisition programs that are intended to have long-term effects on the marketplace. For example, if the goal of the evaluation is to assess cost-effectiveness for stakeholders or regulators, excluding the measurement of market effects in a resource acquisition program could result in under- or overestimating a program’s overall benefits or cost-effectiveness.

While this document focuses on impact evaluation, the three types of evaluation are not mutually exclusive and there are benefits to undertaking more than one type at a time. Process evaluation and market effects evaluation often end up explicitly or implicitly bundled with impact evaluation.

In addition, evaluations often include cost-effectiveness analyses that document the relationship between the value of program results (i.e., energy, demand, and emission savings) and the costs incurred to achieve those benefits. Cost-effectiveness (sometime called cost-benefit) analyses are typically seen as an extension of impact evaluations, but may also take into account market evaluation results considering market penetration over the expected lifetime of the measures. Appendix C includes a brief discussion of cost-effectiveness analyses.

Measurement and verification (M&V) is another term often used when discussing analyses of energy efficiency activities. M&V refers to data collection, monitoring, and analysis used to calculate gross energy and demand savings from *individual sites or projects*. M&V can be a subset of program impact evaluation. Generally speaking, the differentiation between evaluation and project M&V is that evaluation is associated with programs and M&V with projects. The term "evaluation, measurement, and verification" (EM&V) is also frequently seen in evaluation literature. EM&V is a catchall acronym for determining both program and project impacts.

2.5 Notes

1. The Action Plan's *Guide to Resource Planning with Energy Efficiency* is a resource for program planning.

2. For a report presenting DR evaluation issues and options for addressing them, see Violette, D., and D. Hungerford (2007). *Developing Protocols to Estimate Load Impacts from Demand Response Programs and Cost Effectiveness Methods—Rulemaking Work in California*. Presented at International Energy Program Evaluation Conference. <http://www.iepec.org>

3: Impact Evaluation Basics

Chapter 3 describes the key elements of an impact evaluation and introduces the approaches used for determining energy savings. It also presents issues of special interest for conducting impact evaluations, including calculating co-benefits and demand savings, determining persistence of savings, characterizing uncertainty, defining appropriate applications of impact evaluations, and determining avoided emissions.

3.1 Impact Evaluation Process

Impact evaluations determine program-specific induced benefits, which include reductions in energy and demand usage (such as kWh, kW, and therms) and avoided air emissions that can be directly attributed to an energy efficiency program. The basic steps in the evaluation process are:

- Setting the evaluation objectives in the context of the program policy objectives.
- Selecting an approach, defining baseline scenarios, and preparing a plan that takes into account the critical issues.
- Comparing energy use and demand before and after the program is implemented to determine energy/demand savings and avoided emissions.
- Reporting the evaluation results and, as appropriate, working with program administrators to implement recommendations for current or future program improvements.

Basic Impact Evaluation Concepts

- Impact evaluations are used for determining directly achieved program benefits (e.g., energy and demand savings, avoided emissions).
- Savings cannot be directly measured, only indirectly determined by comparing energy use and demand after a program is implemented to what they would have been had the program not been implemented (i.e., the baseline).
- Successful evaluations harmonize the costs incurred with the value of the information received—that is, they appropriately balance risk management, uncertainty, and cost considerations.

The program evaluation process begins with defining and assessing the evaluation objectives. Well-defined objectives indicate what information is needed and the value of that information. The evaluation planning process then indicates the scope and scale of effort required for meeting the objectives (i.e., the cost of obtaining the desired information). The key to successful evaluation is the subsequent comparison of the costs of evaluation with the value of the information received, possibly through an iterative planning process that balances cost and value. Perhaps these two quotes attributed to Albert Einstein best capture the essence of conducting evaluations:

- "Everything should be as simple as it is, but not simpler."
- "Everything that can be counted does not necessarily count; everything that counts cannot necessarily be counted."

3.2 How Energy and Demand Savings Are Determined

The third of the basic steps outlined above has four core components:

1. Gross program energy and demand savings are determined.

2. Gross program savings are converted to net energy and demand savings using a range of possible considerations (e.g., free rider and spillover corrections).[1]
3. Avoided emissions are calculated based on net energy savings.
4. Additional co-benefits are calculated as appropriate. (Typically, the determination of whether to quantify co-benefits is a policy decision.)

Depending on program objectives, it may be desirable to calculate only gross savings. This is done when the only desired result is an estimate of the savings for each project participating in a program—for example, a performance contract to install energy efficiency measures in facilities where the only goal is energy savings. Other instances when only gross savings are calculated is when a predetermined net-to-gross ratio is applied to the results by an overseeing body (such as a regulatory commission) or if producing reliable net savings estimates is simply too expensive or complex.[2] Net savings, in contrast, are calculated when it is of interest to know what savings resulted from the program's influence on program participants and non-participants. This is usually the case when public or ratepayer monies fund the evaluation program or when true (i.e., additional) avoided emission estimates are desired.

As discussed in Section 3.9, the definition of net energy savings used for an energy program sometimes differs from the net energy savings definition used for determining avoided emissions. Thus, while this Guide is organized according to the four steps listed above, each user is free to go as "far down" through the steps as they deem appropriate for their programs and as required to reliably deliver the needed information.

The list of the four steps above does not indicate a time frame for the evaluation activities or reporting. Typically, evaluations are formally organized around annual reporting cycles so that the above steps can be seen as a yearly process. While a year is probably the shortest realistic time frame for reporting complete evaluation results, some entities do provide interim results (such as unverified savings data) on a monthly, quarterly, or semi-annual basis. After the first year's evaluation, the analysis is sometimes referred to as a savings persistence evaluation (see Section 3.5).

Quality Assurance Guidelines

The impact evaluation approaches described in this Guide are based on new and unique analysis of energy and demand savings. Sometimes, however, there is documentation that indicates energy and demand savings were calculated independently of the subject impact evaluation. Although such documentation was not necessarily prepared per pre-determined evaluation requirements, it may be sufficient for meeting the evaluation objectives. Using existing documentation in combination with quality assurance guidelines (QAG) can save significant costs for the program sponsor—and perhaps encourage participation in the program if a portion of evaluation costs are borne by the participants. Essentially, a QAG can help determine whether indicated savings, and the assumptions and rigor used to prepare the documentation, can be used in place of a new evaluation effort.

Gross impact savings are determined using one of, or a combination of, the three different approaches summarized in Section 3.2.1. All of these involve comparing energy usage and demand after the program is implemented to baseline energy use and demand. Net impact savings are determined using one or a combination of four approaches, which are summarized in Section 3.2.2. The approaches used for net and gross savings calculations depends on the objectives of the program, the type of program, and the data and resources available. Selection criteria are discussed in subsequent chapters.

Avoided emissions can be calculated using a variety of approaches that involve determining what emissions are associated with the net energy savings. The definition of net energy savings for an avoided emissions program—along with the sources of emission factors—is discussed in Section 3.9 and Chapter 6.

Other co-benefits of efficiency programs, such as job gain or energy security, are calculated using methods that range from highly rigorous computer models to a simple assessment of anecdotal information. A discussion of co-benefits is included as Section 3.4.

Figure 3-1 summarizes this general approach to the evaluation process.

Evaluation Planning Issues

Chapter 7 of this Guide discusses seven key planning issues to help define policy-specific program evaluation requirements. These are:

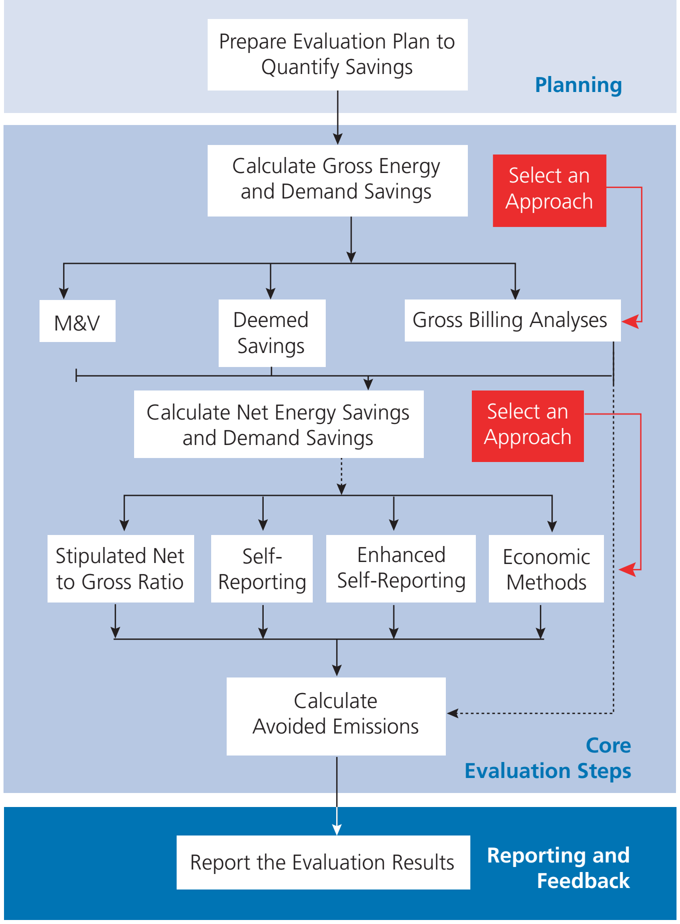
1. Defining evaluation goals and scale, including deciding which program benefits to evaluate.
2. Setting the time frame for the evaluation and reporting expectations.
3. Setting a spatial boundary for evaluation (i.e., what energy uses, emission sources, etc., will be included in the analyses).
4. Defining program baseline, baseline adjustments, and data collection requirements.
5. Establishing a budget in the context of expectations for the quality of reported results.
6. Selecting impact evaluation approaches for gross and net savings calculations and avoided emissions calculations.
7. Selecting who (or which type of organization) will conduct the evaluation.

3.2.1 Approaches for Calculating Gross Energy and Demand Savings

Gross impact savings are determined using one of the following approaches:

- **Measurement and verification (M&V).** A representative sample of projects in the program is selected and the savings from those selected projects are determined and applied to the entire population of projects, i.e. the program. The individual project savings are determined using one or more of the four M&V options defined in the IPMVP (see below). This is the most common approach used for programs involving non-residential facilities, retrofit, or new construction, in which a wide variety of factors determine savings and when individual facility savings values are desired.

Figure 3-1. The Impact Evaluation Process



- **Deemed savings.** Savings are based on stipulated values, which come from historical savings values of typical projects. As with the M&V approach, the savings determined for a sample of projects are applied to all the projects in the program. However, with the use of deemed savings there are no or very limited measurement activities and only the installation and operation of measures is verified. This approach is only valid for projects with fixed operating conditions

Estimation of Gross Energy Savings

The gross energy impact is the change in energy consumption and demand that results directly from program-related actions taken by energy consumers that are exposed to the program, regardless of the extent or nature of program influence on these actions. This is the physical change in energy use after taking into account factors beyond the customer or sponsor's control (e.g., weather). Estimates of gross energy impacts always involve a comparison of changes in energy use over time among customers who installed measures and some baseline level of usage. Baselines may be developed from energy use measurements in comparable facilities, codes and standards, direct observation of conditions in buildings not addressed by the program, or facility conditions prior to program participation.

Estimation of Net Energy Savings

The net energy impact is that percentage of gross energy impact attributable to the program. Estimating net energy impacts typically involves assessing free ridership and spillover, although this Guide discusses additional considerations. "Free ridership" refers to the portion of energy savings that participants would have achieved in the absence of the program through their own initiatives and expenditures. "Spillover" refers to the program-induced adoption of measures by non-participants and participants who did not claim financial or technical assistance for additional installations of measures supported by the program. Other considerations that can be evaluated include the "rebound" or "snapback" effect, transmission and distribution losses (for grid-connected electricity projects) and broader issues such as energy prices and economic conditions that affect production levels. For programs in which participation is not well defined, the concepts of free ridership and spillover are less useful. Estimating net energy impacts for these kinds of programs generally requires the analysis of sales or market share data in order to estimate net levels of measure adoption.

M&V vs. Deemed Savings

For simpler efficiency measures whose performance characteristics and use conditions are well known and consistent, a deemed savings approach may be appropriate. Since they are stipulated and, by agreement, fixed during the term of the evaluation, deemed savings can help alleviate some of the guesswork in program planning and design. However, deemed savings can result in over- or underestimates of savings if the projects or products do not perform as expected—for example, if the energy-efficient lights fail earlier than expected.

Measurement-based approaches are more appropriate for larger and more complex efficiency projects (i.e., those with a significant amount of savings, or "risky" savings). Measured savings approaches are more rigorous than deemed savings approaches and involve end-use metering, billing regression analysis, and/or computer simulation. These approaches add to evaluation costs but may provide more accurate savings values.

Also, deemed savings can be used together with some monitoring of one or two key parameters in an engineering calculation—for example, in a high-efficiency motor program, actual operating hours could be monitored over a full work cycle. This approach is consistent with IPMVP Option A, which is described below.

and well-known, documented stipulation values (e.g., energy-efficient appliances such as washing machines, computer equipment and refrigerators, lighting retrofit projects with well understood operating hours). This approach involves multiplying the number of installed measures by the estimated (or deemed) savings per measure. Deemed savings values are only valid when they are derived from documented and validated sources, such as historical evaluations, and only apply to the most common efficiency measures. Deemed savings are the per-unit energy savings values that can be claimed from installing specific measures under specific operating situations. Examples include agreed-upon savings per fixture for lighting retrofits in office buildings, with specific values for lights in private offices, common areas, hallways, etc.

- **Large-scale data analysis.** Statistical analyses are conducted on the energy usage data (typically collected from the meter data reported on utility bills) for all or most of the participants and possibly non-participants in the program. This approach is primarily used for residential programs with relatively homogenous participants and measures, when project-specific analyses are not required or practical.

3.2.2 Approaches for Calculating Net Energy and Demand Savings

The difference between net and gross savings is specified as a net-to-gross ratio (NTGR). The four approaches for determining the NTGR are:

- **Self-reporting surveys.** Information is reported by participants and non-participants, without independent verification or review.

- **Enhanced self-reporting surveys.** The self-reporting surveys are combined with interviews and independent documentation review and analysis. They may also include analysis of market-based sales data.

- **Econometric methods.** Econometrics is the application of statistical tools and techniques to economic issues and economic data. In the context of calculating net energy savings, statistical models are used to compare participant and non-participant energy and demand patterns. These models often include survey inputs and other non-program-related factors such as weather and energy costs (rates).

- **Deemed net-to-gross ratios.** An NTGR is estimated using information available from evaluation of similar programs. This approach is sometimes used by regulatory authorities.

It is not unusual for combinations of these approaches to be used. For example, rigorous econometric methods may be used every three years with self-reported or deemed NTGRs used for the other program years. If a previous econometric study is considered more reliable, its results may be used as the deemed value. Another option is to calibrate self-reported calculations to align with the previous study's results.

National Grid Net Savings Example

In 2006, National Grid undertook a study of free ridership and spillover in its commercial and industrial energy efficiency programs. That study identified a free ridership rate of 10 percent and a spillover rate of 14 percent for custom installations as determined using the Design 2000*plus* software program. The net-to-gross ratio for custom installations is equal to:

$$\begin{aligned} NTGR &= (1 - \text{free ridership} + \text{spillover}) \\ &= (1 - 0.10 + 0.14) \\ &= 1.04 \end{aligned}$$

In this case, net savings for custom installations in National Grid's Design 2000*plus* Program are 4 percent higher than gross savings.

Provided by National Grid based on a report from PA Consulting Group, 2006.

Note that gross energy savings may be determined and reported on a project-by-project or program-wide basis. Net savings can also be determined on a project-by-project or program-wide basis, but are almost always only reported on a program-wide basis. This program-wide reporting is done in terms of the NTGR. For example, a NTGR of 90 percent would indicate that, on average, 90 percent of the indicated gross savings are attributed to the influences of the program.

Lastly, the net savings approaches described here work best in regions with new program efforts. In regions with a long history of program efforts, such approaches may understate a program's effects because of the program's long-term influences and the difficulty of separating out one program's influences from other influences.

3.3 Calculating Demand Savings

For efficiency programs, determining energy savings is almost always a goal of impact evaluations. A program's electrical demand savings are also often of interest, and for some programs are a primary goal.[3] Energy usage and savings are expressed in terms of consumption over a set time-period and are fairly straightforward to define (e.g., therms of natural gas consumed per month, MWh of electricity consumed over a year, season, or month, etc). Energy savings results may also be reported by costing period, which break the year into several periods coinciding with a utility rate schedule. Examples include peak and off-peak periods or summer and winter periods.

Demand savings are expressed in terms of kW or MW, which indicate *rates* of consumption. Historically, demand savings (particularly peak demand savings rather than simple annual average demand savings) have been much harder to define and determine than energy savings. This is because determining demand savings requires data collecting and analysis for specific time periods—for example, data might be required for summer weekdays between noon and 6 p.m., as compared to aggregated monthly utility meter data. However, with technology advances lowering the cost of meters, sophisticated wired and wireless sensors, and the related software and increasing availability and use of utility "smart" meters that collect real time data, it is becoming easier to cost-effectively collect the data needed to calculate demand savings.

Regional Coincident Peak Demand

Coincident peak demand can be considered for a region as well as for a single utility. For example, in New England, utilities are interested in looking at demand savings coincident with the ISO-New England peak, which is defined for both the summer and for the winter. The individual utilities' peaks may or may not be at the same time.

Examples of demand savings definitions are:

- **Annual average demand savings**—total annual energy savings divided by the hours in the year (8,760). In the Northwest United States, this is termed average MW, or MWa.
- **Peak demand reductions**—there are several definitions in use for peak demand reduction. They all involve determining the maximum amount of demand reduction during a period of time, whether that be annual, seasonal, or a specific period such as during summer weekday afternoons or during winter peak billing period hours. If peak demand reduction is to be reported as part of an evaluation, the term must be clearly defined.
- **Coincident peak demand reduction**—the demand savings that occur when the servicing utility is at its peak demand from all (or segments) of its customers. This indicates how much of a utility's peak demand is reduced during the highest periods of electricity consumption. Calculating coincident peak demand requires knowing when the utility has its peak (which is not known until the peak season is over). A term used to describe the relationship of facility electrical loads to coincident peak demand is "diversity factor": the ratio of the sum of the demands of a group of users to their coincident maximum demand, always equal to or greater than 1.0.
- **Demand response peak demand reduction**—for demand reduction programs, it is desired to know what reduction occurs when there is a call for demand reductions. The evaluation can be of the

(a) level of demand reduction that has been pledged or enabled through testing and inspection or (b) level of demand reduction that has been achieved using a variety of methods, some of which are included in this Guide and some of which are specific to demand response.

The calculation for demand savings is straightforward:

demand savings = energy savings ÷ time period of energy savings

Each of the gross impact evaluation approaches, to varying degrees of accuracy and with varying degrees of effort, can be used to determine demand savings using the above equation. The "trick" is collecting the energy savings data for the intervals of interest (i.e., the time period in the above equation). If annual average demand savings are the only data required, then only annual energy savings data are necessary. However, if peak demand reduction, coincident demand reduction, or demand response peak demand reduction values are desired, then hourly or 15-minute energy savings data, or estimates, are required.

Ideally, evaluation results would indicate 8,760 hours per year of energy savings data that could be easily translated into hourly demand savings. In practice there are both primary and secondary methods for determining demand savings. Primary methods involve collecting hourly or 15-minute demand data during the periods of interest, for example during the peak hours of the summer months (peak season) of each year.

Sources of hourly or 15-minute data include facility interval-metered data, time-of-use consumption billing data, monthly billing demand data, and field-measured data. When interval or time-of-use consumption data are available, they can be used for regression analysis to account for the effects of weather, day type, occupancy, and other pertinent change variables on demand savings. Of course, hourly demand data can require hourly independent variable data (e.g., weather) for proper regression analysis.

Secondary methods rely upon collected energy consumption data that are only available as averaged values over longer periods, such as monthly or even annually. When longer periods are used, demand impacts can also be estimated from energy impacts by applying a series of standard *load shapes* to allocate energy consumption into costing period bins. These load shapes (for whole facilities or by end-use) may be available from other studies for related programs in similar

Demand Response Evaluation

Demand response (DR) programs are specifically aimed at reducing peak demand, and some of the concepts and principles discussed in this Guide can be used for DR program evaluation.[5] Protocols for DR evaluation are under development in California and are currently under review and comment (available at <http://www.cpuc.ca.gov/static/hottopics/1energy/draftdrloadimpactprotocols.doc>). Several studies of DR impacts in eastern U.S. markets have also been conducted in recent years that deploy complex econometric price modeling and simulation to estimate baselines (see, for instance, LBNL studies at <http://eetd.lbl.gov/ea/EMP/drlm-pubs.html>). The draft California DR protocols identify numerous issues relating to evaluation of DR that are not addressed in energy efficiency evaluation protocols because they do not apply to efficiency. These include the difference in estimating impacts from event versus non-event programs, estimating program-wide impacts (for resource planning) versus customer-specific impacts (for settlement), and representative-day versus regression baseline estimation.

In 2007, the Independent System Operator of New England (ISO-NE) developed an M&V manual that describes the minimum requirements the sponsor of a demand resource project must satisfy to qualify as a capacity resource in New England's wholesale electricity Forward Capacity Market. A text box in Chapter 7 describes the EM&V requirements developed for that program. The ISO-NE EM&V requirements can be found at <http://www.iso-ne.com/rules_proceds/isone_mnls/index.html>.

markets. One source for the data is the energy savings load shapes, by measure, that are included in the California Database for Energy Efficiency Resources (DEER).[4]

3.4 Co-Benefits

This Guide describes techniques for documenting three categories of impacts or benefits associated with energy efficiency programs: energy savings, demand savings, and avoided air emissions. However, there are other potential benefits of energy efficiency. These include:

- Avoided transmission and distribution capital costs and line losses.
- Reliability net benefits.
- Voltage support and power quality benefits.
- Environmental net benefits (in addition to air pollution and climate impacts, the most common considerations relate to water).
- Energy price effects.
- Economic impacts (e.g., employment, income, trade balances, tax revenues).
- National security impacts.

An important category of "co-benefits" is participant non-energy benefits (NEBs). Participant NEBs can include non-market goods, such as comfort and safety, as well as water savings and reduced operation and maintenance costs. Other possible positive NEBs include reduced eyestrain due to improved lighting quality and higher resale value associated with energy-efficient building upgrades. However, non-energy benefits can also be negative. Examples of negative NEBs are aesthetic issues associated with compact fluorescent bulbs and increased maintenance costs due to unfamiliarity with new energy-efficient equipment.

Often, such co-benefits are listed but not quantified. This is because of the lack of standardized and agreed-upon methods for quantifying these benefits, the cost of doing such quantification, and the sense that the majority of financial benefits are associated with saved energy costs.

Evaluating Participant Non-Energy Benefits

NEBs can be evaluated through a range of survey approaches:

- Contingent valuation (CV) survey techniques directly ask respondents' willingness to pay for a particular good.
- Direct query (DQ) approaches ask respondents to value NEBs relative to a given parameter, such as the energy savings achieved on their project. To assist respondents, these surveys often use a scale or provide the dollar value of the energy savings.
- Conjoint analysis (CA) survey techniques provide respondents with descriptions of different scenarios or levels of NEBs, asking them to either rank or choose between the different options presented. Econometric techniques are then applied to calculate the "utility" or relative value of each attribute.

All of these approaches have benefits and drawbacks. The industry standard, to date, has been CV and DQ approaches. However, in recent years, NYSERDA has pioneered the joint use of DQ and CA survey methods on its New York Energy $mart Program. Thus far, the DQ and CA approaches have resulted in individual NEB values within the same general range (note that NYSERDA uses the term "non-energy indicators"). However, values derived by CA fall toward the lower end of the range. This could be due to many factors, not the least of which is the more limited set of non-energy co-benefits that can reasonably be covered in CA surveys.

Source: NYSERDA, 2006.

Table 3-1. Wisconsin Focus on Energy Value of Non-Energy Benefits by Program Area

July 1, 2001–June 30, 2007

<table>
<tr><th rowspan="2">Program Area</th><th colspan="2">Value of Non-Energy Benefits*</th></tr>
<tr><th>FY07
as of June 30, 2007</th><th>Program to Date
as of June 30, 2007</th></tr>
<tr><td>Business Programs</td><td>$4.4 million</td><td>$17.9 million</td></tr>
<tr><td colspan="3"><i>Example Benefits from Business Programs:</i>
• Reduced equipment maintenance
• Increased employee morale
• Increased equipment life
• Increased productivity
• Reduced waste generation
• Reduced defects and errors
• Increased sales
• Reduced non-energy costs
• Reduced personnel needs
• Reduced injuries and illnesses</td></tr>
<tr><td>Residential Programs</td><td>$4.5 million</td><td>$34.6 million</td></tr>
<tr><td colspan="3"><i>Example Benefits from Residential Programs:</i>
• Increased safety resulting from a reduction of gases such as carbon monoxide due to the installation of a new high-efficiency furnace
• Fewer illnesses resulting from elimination of mold problems due to proper air sealing, insulating and ventilation of a home
• Reduced repair and maintenance expense due to having newer, higher quality equipment
• Increased property values resulting from installation of new equipment
• Reduced water and sewer bill from installation of a horizontal-axis washing machine, which uses much less water than a conventional washing machine</td></tr>
<tr><td>Renewable Energy Programs</td><td>$163,128</td><td>$726,902</td></tr>
<tr><td colspan="3"><i>Example Benefits from Renewable Energy Programs:</i>
• Greater diversity of primary in-state energy supplies
• Use of wastes as a fuel instead of disposal
• Increased ability to handle energy emergencies or generation shortfalls
• Increased sales of renewable energy byproducts</td></tr>
</table>

*Method of applying value is under review.

Source: TecMarket Works, 2002, 2003; State of Wisconsin, 2007.

However, cost-effectiveness analysis requires that *at least* the most important types of benefits and costs be valued in dollar terms. This "monetization" of benefits and costs is necessary in order to facilitate the comparison of benefits and costs and to allow the determination of whether benefits outweigh costs. Of course, not all program impacts may be amenable to valuation; nonetheless, program selection and continuation decisions are greatly facilitated to the extent that such valuation can be accomplished, and therefore at least a listing of the non-quantified co-benefits is commonly included in evaluation reports.

In summary, including non-energy co-benefits in the evaluation process tends to increase the value of saved energy and both justify additional energy efficiency investment and demonstrate the cost-effectiveness of more aggressive efficiency activities, as compared to supply side investments.

New York and Wisconsin are two states, among others such as California and Massachusetts, that estimate co-benefits in their evaluations:

- NYSERDA undertakes a macroeconomic impact analysis of the New York Energy $mart Program by comparing the impacts of program expenditures and energy savings to a basecase estimate of the impacts of the system benefits charge (SBC) programs. The basecase is the impact that SBC funds would have had on the New York economy had they been retained by participating utility customers in the absence of the program. The program case estimates the impact on the New York economy of SBC funds allocated to the portfolio of New York Energy $mart Program expenditures. The net macroeconomic impacts are expressed in terms of annual employment, labor income, total industry output, and value added.
- Table 3-1, from a Wisconsin Focus on Energy report illustrates the state's evaluation of energy efficiency co-benefits (TecMarket Works, 2002, 2003, 2005).

3.5 Persistence

One important evaluation issue is how long energy savings are expected to last (persist) once an energy efficiency activity has taken place. A persistence study measures changes in the net impacts over time. These changes are primarily due to retention and performance degradation, although in some instances changes in codes or standards or the impact of "market progression"[6] can also reduce net savings. Effective useful life (EUL) is a term often used to describe persistence. EUL is an estimate of the median number of years that the measures installed (or activities implemented) under a program are still in place and operable.

Persistence studies can be expensive undertakings. Past experience indicates that long periods of time are needed for these studies, so that large samples of failures are available and technology failure and removal rates can be better documented and used to make more accurate assessments of failure rates. The selection of what to measure, when the measurements should be launched, and how often they should be conducted is a critical study planning consideration (CPUC, 2006).

Note also that the energy savings achieved over time is a difference rather than a straight measurement of the program equipment or a consumer behavior. For example, the efficiency of both standard and high-efficiency equipment often decreases over time; thus, savings are the difference over time between the energy usage of the efficient equipment/behavior and the standard equipment/behavior it replaced.

The basic approaches for assessing persistence are:

- Use of historical and documented persistence data, such as manufacturer's studies or studies done by industry organizations such as ASHRAE.
- Laboratory and field testing of the performance of energy-efficient and baseline equipment.
- Field inspections, over multiple years, of efficiency activities that constitute a program.

- Non-site methods such as telephone surveys and interviews, analysis of consumption data, or use of other data (e.g., data from a facility's energy management system).

The California Evaluation Protocols contain a complete section on persistence analyses and can be used to learn more about this subject.

3.6 Uncertainty

Perhaps the greatest challenge in evaluating energy efficiency programs is the impossibility of direct measurement of the primary end result—energy savings. Energy savings are the reduction from a level of energy use that did not happen. What can be measured is actual energy consumption after, and sometimes before, the energy efficiency actions. Consequently, the difference between (a) actual energy consumption and (b) what energy consumption would have been had the efficiency measures not been installed is an *estimate* of energy (and demand) savings.

Since program evaluations seek to reliably determine energy and demand savings with reasonable accuracy, the value of the estimates as a basis for decision-making can be called into question if the sources and level of uncertainty of reported savings estimates are not fully understood and described. While additional investment in the estimation process can reduce uncertainty, tradeoffs between evaluation costs and reductions in uncertainty are inevitably required.

Thus evaluation results, like any estimate, are reported as expected values including some level of variability (i.e., uncertainty). Uncertainty of savings level estimates is the result of two types of errors, systematic and random.

1. **Systematic errors** are those that are subject to decisions and procedures developed by the evaluator and are not subject to "chance." These include:
    - Measurement errors, arising from meter inaccuracy or errors in recording an evaluator's observations.
    - Non-coverage errors, which occur when the evaluator's choice of a sampling frame excludes part of the population.
    - Non-response errors, which occur when some refuse to participate in the data collection effort.
    - Modeling errors, due to the evaluator's selection of models and adjustments to the data to take into account differences between the baseline and the test period.
2. **Random errors**, those occurring by chance, arise due to sampling rather than taking a census of the population. In other words, even if the systematic errors are all negligible, the fact that only a portion of the population is measured will lead to some amount of error. Random errors are sometime called sampling errors.

The distinction between systematic and random sources of error is important because different procedures are required to identify and mitigate each. The amount of random error can be estimated using standard statistical tools, while the systematic errors discussed above cannot be easily estimated. In most instances, evaluators simply try (within budget limitations) to prevent systematic errors from occurring. Thus, uncertainty is typically calculated through the consideration of random errors.

Assuming that a random procedure has been used to select the sample, sampling error can be estimated by using the laws of probability and sampling distributions. In other words, the potential magnitude of the sampling error for any value calculated from a sample can usually be estimated. The common factors for reporting sampling uncertainty are *confidence* and *precision*. Confidence is the likelihood that the evaluation has captured the true impacts of the program within a certain range of values, with this range of values defined as precision. (For additional information on calculating uncertainty, see ASHRAE, 2002, and WRI and WBCSD, 2005a.)

Sampling can be a particularly important aspect of an evaluation design, and decisions about the sample size are one of the key influences on the overall uncertainty of the evaluation. Evaluators typically do not have access to an entire population of interest (e.g., all small commercial customers participating in a program), either because the population is too large or the measurement process is too expensive or time-consuming to allow more than a small segment of the population to be observed. As a result, they must base their decisions about a population on a small amount of sample data. Examples of impact evaluation samples are:

- **Residential efficiency retrofit program**—a sample of homes is selected for analysis (versus all of the homes that were retrofitted). The sample may be organized into homes with similar physical characteristics, similar occupants, similar vintages, etc.
- **Commercial building lighting retrofit program**—a sample of the "spaces" (offices, hallways, common areas, etc.) is selected for inspection, metering, and analysis from different buildings that participated in the program.
- **Industrial motors retrofit program**—a sample of motors that were installed is selected for metering of power draw during a range of operating conditions and time periods.
- **New construction building incentive program**—all of the buildings in a program are selected for analysis but only within a certain time period (e.g., one month per year).
- **NTGR analysis of participants in an efficiency program**—a sample of participants and a sample of non-participants are selected for interviews.

Evaluation of savings uncertainty is an ongoing process that can consume time and resources. It also requires the services of evaluation contractors who are familiar with data collection and analysis techniques. And, of course, reducing errors usually increases evaluation cost. Thus, the need for reduced uncertainty should be justified by the value of the improved information. That is, is the value worth the extra cost?

Appendix D briefly presents some statistical fundamentals that are important for any discussion of uncertainty, with an emphasis on sampling issues. These issues apply to energy, demand, and non-energy benefit evaluations. Appendix D is not intended as a primer on statistics, but to give program and evaluation managers and regulators some basic information from which they can specify what they expect their evaluation contractors to address with respect to uncertainty and sampling in evaluation plans and reports. Its purpose is to provide an overview of (a) the range of factors that contribute to uncertainty, (b) an understanding of how each factor contributes to uncertainty and why it is important to assess its impact on uncertainty, and (c) an awareness of what steps can be taken to reduce the level of uncertainty in evaluation results.

3.7 Appropriate Applications of Impact Evaluations

It is appropriate to conduct impact evaluations when the evaluation objectives are to:

- Determine, quantify, and document energy and demand savings and avoided emissions that *directly* result from an efficiency program,
- Document the cost-effectiveness of an efficiency program, or
- Inform current or future program implementers of the savings actually achieved from particular measures or program strategies.

Producing savings directly means that the link between the program activity and the savings is clear, straightforward, and relatively fast. Programs based on information, education, marketing, promotion, outreach, and similar efforts do not provide such direct impacts. For these programs there can be a more tenuous link between the program activities and any eventual savings. Savings obtained from these programs depend upon inducing some form of behavior change (such as turning off lights, independently purchasing and installing efficient equipment, or participating in a more direct efficiency program). Thus, if the primary objective of a program is providing savings *indirectly* (such as through a market transformation program), then the primary evaluation effort will most likely be a market effects evaluation, not an impact evaluation (though an impact evaluation could still be conducted to quantify any direct savings). This may be particularly true when there are overlapping programs, such as an education program working in tandem with a resource acquisition program to convince customers to participate (through education) and then actually incents their participation through rebates (i.e., resource acquisition).

Cost-effectiveness assessments require information on quantified gross or net savings. Thus, in order to calculate cost-effectiveness, an impact evaluation must be conducted—if overall market costs and savings are to be included in the analysis, a market effects study is also required. The costs and savings, possibly including avoided emissions, are then monetized and compared to determine cost-benefit indicators. In terms of program objectives, the oversight and feedback inherent to evaluation also helps maintain cost-effectiveness.

3.8 Evaluation Characteristics and Ethics

Evaluation processes are defined by the following principles.

- **Completeness and transparency.** Results and calculations are coherently and completely compiled. Calculations are well documented in a transparent manner, with reported levels of uncertainty that allow verification by an independent party. The scope of the documentation takes into account the relevant independent variables that determine benefits and the baseline is properly defined. In addition, documentation and reporting include all relevant information in a coherent and factual manner that allows reviewers to judge the quality of the data and results. Among the key qualities of a good, transparent analysis are:
    - Project descriptions indicate the activity and the variables determining energy savings.
    - Critical assumptions are stated and documented.
    - Documentation is presented in a format that allows the reviewer to follow a connected path from assumptions to data collection, data analysis, and results.
    - Levels and sources of uncertainty are reported.
- **Relevance and balance in risk management, uncertainty, and costs.** The data, methods, and assumptions are appropriate for the evaluated program. The level of effort expended in the evaluation process is balanced with respect to the value of the savings (and avoided emissions), the uncertainty of their magnitude, and the risk of over- or underestimated savings levels. Benefits are calculated at a level of uncertainty such that the savings are neither intentionally over- nor underestimated and the quality of the reported information is sufficient for maintaining the integrity of the program being evaluated.
- **Consistency.** Evaluators working with the same data and using the same methods and assumptions will reach the same conclusions. In addition, for efficiency programs that are part of broad efforts, such as utility resource procurement programs or emissions cap and trade systems, energy and demand savings and avoided emissions calculated from one program are as valid as those generated from any other actions, whether demand-side or supply-side. This allows for comparison of the range of energy resources, including energy efficiency. Examples of consistency include:
    - Using the same measurement techniques for determining the baseline and reporting period electricity consumption of a system.
    - Using the same assumptions for weather, indoor environment (e.g., temperature set points, illumination levels, etc.), and occupancy in a building for baseline and reporting period energy analyses.

Another characteristic that is cited, particularly in the GHG emissions evaluation literature, is conservativeness. With counterfactual baselines, uncertainty is inherent and savings estimates are prone to a certain degree of subjectivity. Because of this subjectivity, and possibly a lack of relevant information, some believe that "conservativeness" should be added to the list of principles for the purpose of counteracting a natural tendency toward savings inflation. There are many real-world incentives for people to over-report savings or avoided emissions, and fewer incentives working the other way. This subjective bias may be difficult to keep in check without an explicit directive to be conservative. However, others believe that credibility, not conservativeness, is the desired characteristic, and that underestimates can be just as biased and damaging as overestimates. Like other evaluation policy decisions, this one is best made by those responsible for defining evaluation objectives.

Related to the characteristics of the evaluation itself, the credibility of evaluators is essential for providing credible findings on the results from the program and for providing recommendations for program refinement and investment decisions. Thus, evaluation ethics are a critical foundation for the activities described in this Guide. The American Evaluation Association (AEA) has a set of guiding ethical principles for evaluators. Located on AEA's Web site <http://www.eval.org>, these principles are summarized here:

- **Systematic inquiry**—evaluators conduct systematic, data-based inquiries.
- **Competence**—evaluators provide competent performance to stakeholders.
- **Integrity/honesty**—evaluators display honesty and integrity in their own behavior, and attempt to ensure the honesty and integrity of the entire evaluation process.
- **Respect for people**—evaluators respect the security, dignity, and self-worth of respondents, program participants, clients, and other evaluation stakeholders.
- **Responsibilities for general and public welfare**—evaluators articulate and take into account the diversity of general and public interests and values that may be related to the evaluation.

3.9 Calculating Avoided Emissions

State and federal policymakers and utility regulators are broadening the scope of evaluation by integrating efficiency programs focused on (a) achieving energy savings with programs that focus on other objectives such as reducing dependency on fossil fuels (e.g., renewable energy and combined heat and power—see Appendix F), (b) reducing the need for investments in generating capacity (demand response), and (c) investing in technologies that help to mitigate pollution and greenhouse gas emissions. Because the avoided emissions benefits of energy efficiency are of particular interest, this section provides a brief overview of efficiency-induced avoided emissions and discusses some specific issues related to avoided emissions calculations: additionality, boundary area definitions, and aspects of cap and trade programs. Chapter 6 builds on this information and provides information on the actual calculation of avoided emissions once the energy savings from an efficiency program have been determined.

3.9.1 Energy Efficiency and Avoided Emissions

Energy efficiency can reduce emissions associated with the production of electricity and thermal energy from fossil fuels. However, historically, emissions reductions from efficiency projects are described only subjectively as a non-quantified benefit. This is changing with increasing interest in quantifying these benefits, both for conventional pollutants such as sulfur dioxide (SO_2), nitrogen oxides (NO_X), mercury (Hg), and particulates (PM) as well as for greenhouse gases (GHGs)—primarily carbon dioxide (CO_2) —from fossil fuel combustion.

Energy efficiency is particularly important for reducing GHGs because there are few options or "controls" for reducing CO_2 emissions from combustion once the CO_2 is formed. The implication is that energy efficiency can be the lowest cost option for reducing GHG emissions. The importance of efficiency also becomes clear in light of the fact that approximately 61 percent of all human-induced or "anthropogenic" GHG emissions come from energy-related activities—the breakout of global energy-related GHG emissions is estimated at 40 percent for electricity and heat, 22 percent for transport, 17 percent for industry, 15 percent for other fuel combustion, and 6 percent for fugitive emissions (Baumert et al., 2005).

For any type of energy efficiency program, the avoided air emissions are determined by comparing the emissions occurring after the program is implemented to an estimate of what the emissions would have been in the absence of the program—that is, emissions under a baseline scenario. Conceptually, avoided emissions are calculated using the net energy savings calculated for a program and one of two different approaches:

1. **Emission factor approach**—multiplying the program's net energy savings by emission factors (e.g., pounds of CO_2 per MWh) representing the characteristics of displaced emission sources to compute hourly, monthly, or annual avoided emission values (e.g., tons of NO_X or CO_2). The basic equation for this approach is:

   $$\text{avoided emissions}_t = (\text{net energy savings})_t \times (\text{emission factor})_t$$

2. **Scenario analysis approach**—calculating a base case of sources' (e.g., power plants connected to the grid) emissions without the efficiency program and comparing that with the emissions of the sources operating with the reduced energy consumption associated with the efficiency program. This is done with sophisticated computer simulation approaches known as "dispatch models" (see Chapter 6). Scenario analysis is typically only used with electricity-saving programs. The basic equation for this approach is:

   $$\text{avoided emissions} = (\text{base case emissions}) - (\text{reporting period emissions})$$

One important consideration for both of these approaches is that the net energy savings calculated for the purposes of an energy resource program may be different from the net savings that need to be calculated to meet the requirements of an avoided emissions program. Three potential causes of the difference are:

1. Different definitions of *additionality*.
2. Different definitions of *boundary areas*.
3. The characteristics of *emissions control mechanisms/regulations* that may be in place.

The first two items are discussed in Sections 3.9.2 and 3.9.3. The "cap and trade" emissions control mechanism and its features with respect to energy efficiency are discussed in Section 3.9.4. Although it is not the only option to achieve widespread emissions reductions, it is addressed here because of its unique characteristics and current popularity. Following these subsections is a brief overview of the possible objectives associated with calculating avoided emissions and how they can affect decisions about what calculation approaches should be used and what specific issues should be addressed.

3.9.2 Additionality

Additionality is the term used in the emission mitigation industry for addressing the key question of whether a project will produce reductions in emissions that are additional to reductions that would have occurred in the absence of the program activity. This is directly related to the efficiency evaluation issue of defining proper baseline conditions and free ridership, as described in Chapters 4 and 5, respectively. As the baseline is a "what-if" value, it cannot be directly measured and must be inferred from available information.

While the basic concept of additionality may be easy to understand, there is no common agreement on the procedures for defining whether individual projects or whole programs are truly additional (i.e., different than a baseline scenario). As such, there is no technically correct level of stringency for additionality rules. Evaluators may need to decide, based on their policy objectives, what tests and level of scrutiny should be applied in additionality testing. For example, program objectives that focus on obtaining avoided emissions credits as part of a regulatory program may necessitate stringent additionality rules. On the other hand, programs that are primarily concerned with maximizing energy efficiency and only need to approximately indicate avoided emissions may establish only moderately stringent rules.

3.9.3 Assessment Boundary Issues: Primary and Secondary Effects/Direct and Indirect Emissions

The "emissions assessment boundary" is used to define and encompass all the energy uses and emission sources affected by activities in a program. (The "assessment boundary" and "primary/secondary" terminology is drawn from WRI and WBCSD, 2005b). For avoided air emissions, the assessment boundary can be much larger than the boundary for calculating energy and demand savings, including changes to emission rates and volumes beyond avoided emissions associated with less energy use at the efficiency project sites.

Direct and indirect emissions are two categories for consideration when setting an emissions assessment boundary. Direct emissions are changes in emissions at the site (controlled by the project sponsor or owner). For efficiency projects affecting onsite fuel use—for example high-efficiency water heaters or boilers, the avoided emissions are direct. Indirect emissions are changes in emissions that occur at a source away from the project site (e.g., a power plant). Indirect emissions are the primary source of avoided emissions for electrical efficiency programs.

When defining the assessment boundary, one must also consider intended and unintended consequences, also called primary and secondary effects.

- A primary effect is the intended change in emissions caused by a program. Efficiency programs generally have only one primary effect—energy savings at facilities that consume energy, translating into avoided emissions.
- A secondary effect is an unintended change in emissions caused by a program. Secondary effects are sometimes called "leakage." Leakage and interactive effects (defined in Chapter 4) are similar concepts, although leakage is a more "global" issue whereas interactive effects tend to be considered within the facility where a project takes place. Two categories of secondary effects are:
    - One-time effects—changes in emissions associated with the construction, installation, and establishment or the decommissioning and termination of the efficiency projects—net of the same level of efficiency activity in the baseline scenario.
    - Upstream and downstream effects—recurring changes in emissions associated with inputs to the project activities (upstream) or products from the project activity (downstream) relative to baseline emissions. For example, one upstream effect of possible concern (however unlikely) for efficiency programs is that if efficiency programs displace energy sales and emissions in one area, the same amount of energy consumption, and related emissions, might be shifted elsewhere.

Secondary effects, *outside the facility where the efficiency project takes place*, are typically minor relative to the primary effects of energy efficiency programs—particularly when compared to baseline secondary effects. For example, the manufacturing, maintenance, and installation of energy-efficient motors have no meaningfully different associated emissions than the emissions associated with standard efficiency motors. In some cases, however, secondary effects can undermine the primary effect; therefore, the emissions assessment boundary should be investigated, even if to only document that there are no secondary effects.

In summary, when evaluating the avoided reductions associated with efficiency programs, it is important to properly define the assessment boundary, and ideally to account for all primary effects (the intended savings) and secondary effects (unintended positive or negative effects) and all direct emissions (at the project site) and indirect emissions (at other sites).

3.9.4 Special Issues for Capped Pollutants Under Cap and Trade Programs

There are numerous mechanisms for controlling pollutants and greenhouse gas emissions, and "cap and trade" is one of them. Under a cap and trade program, an overall emission tonnage cap is set for an affected sector or set of plants. Allowances are created to represent the emission of each unit (e.g., one ton) of pollution under the allowable cap. The primary compliance requirement is that each plant must hold allowances equal to its actual emissions at the end of each compliance period. However, there is no fixed emission cap or limit on an individual plant and each plant's emissions are not limited to the allowances that it initially receives or buys at auction (depending on how allowances are allocated). It may purchase additional allowances from another plant or sell allowances if it has a surplus.

Examples of cap and trade programs in the United States are:

- The Title IV acid rain SO_2 trading program sets a cap on annual SO_2 emissions for U.S. power plants.
- NO_X emissions are currently capped during the summer for 21 eastern states and will be capped year-round starting in 2009 for most of the eastern United States plus Texas.
- CO_2 emissions will be capped in the 10 states of the Northeastern Regional Greenhouse Gas Initiative starting in 2009, California has enacted legislation to limit GHG emissions, the Western Regional Climate Action Initiative may adopt a cap, and other states are working on similar programs.

The level of the cap is an important aspect of a cap and trade program. Emissions may not exceed the cap, and they are also unlikely to be below the cap over any substantial time period. The reason for this is that a unit that emits fewer allowances than it has available may sell those allowances to another unit, which will then use them to pollute. Plants may also "bank" unused allowances to use in a future year. Thus, the overall sector will always emit approximately at the cap level.

The fact that capped emissions tend to remain at the cap level is very relevant to the effect of energy efficiency. When emissions are not capped, energy efficiency reduces the output of electricity generators and thus reduces emissions. As noted, this is not typically true for emissions from sources subject to caps (e.g., large boilers, power plants). Reductions in these capped emissions make extra allowances available for other entities to use. This means that these "efficiency" allowances can be sold in the market and used elsewhere or banked for use in a later year, such that total emissions will remain roughly equal to the cap level.

There are, however, mechanisms by which efficiency programs under a cap and trade system can claim avoided emissions. These are that (a) the "efficiency allowances" are retired (removed from the market) or (b) policies are put in place to ensure that the emissions trading cap and the number of allowances allocated are reduced commensurate with the prevailing level of energy efficiency. Since the goal of the trading program is typically not to go below the cap but to achieve the cap at the lowest possible cost to society, energy efficiency contributes to the primary goal of the cap and trade program by helping to minimize compliance costs. In addition, efficiency programs may reduce emissions from non-capped emission sources and directly claim avoided emissions if properly calculated.

Another way for energy efficiency programs to create actual reductions under a cap and trade program is to assign allowances to the efficiency activities and retire them. For example, some states have created special set-aside allocations of allowances in their NO_X trading programs for energy efficiency projects (see <http://www.epa.gov/cleanenergy/pdf/eere_rpt.pdf>). Qualified project sponsors that obtain these allowances can choose to retire them to make emissions reduction claims and avoid the expense of an allowance purchase that would otherwise be necessary to make such claims. However, sponsors may also sell the allowances to finance the efficiency project, in which case they may not claim the reduction. The U.S. EPA has developed EM&V guidance for the NO_X set-aside program covering avoided emissions calculations for both renewables and efficiency projects (see <http://www.epa.gov/cleanenergy/pdf/ee-re_set-asides_vol3.pdf>).

3.9.5 Avoided Emissions Calculations for Different Objectives

Avoided emissions calculations have a wide range of specific applications, such as voluntary and mandatory GHG offset programs and NO_X cap and trade programs with energy efficiency allowance set-asides. These programs have varying requirements for what are considered legitimate avoided energy emissions. Those interested in creating tradable offsets, allowances, or other program-specific credits should consult the regulations of the specific program they are interested in with respect to additionality and boundary area definitions, as well as other issues specific to the given program.

However, the following are some rule-of-thumb recommendations based on what the objective is for calculating the avoided emissions:

- **Calculating avoided emissions primarily for informational purposes.** When the primary goal of an efficiency program is saving energy or demand, the avoided emissions are often reported only to subjectively and approximately indicate a co-benefit. In this situation, the expectations for the certainty of the avoided emission values are not high and the avoided emission estimates are not used in a regulatory or market scheme where a monetary value is ascribed to the avoided emissions. In this situation, a simple approach as described in Chapter 6 can be appropriate. It is typical that (a) additionality is simply assumed, (b) emissions boundary area issues are ignored, and (c) the energy savings are simply those reported for the program, whether net or gross. These savings are then multiplied by appropriate, preferably time-dependent, emission factors to calculate avoided emissions. With this type of calculation, the uncertainty of the avoided emissions estimate is probably high. As noted above, there may not even be actual avoided emissions if the efficiency activities reduce emissions from capped sources regulated under a cap and trade program.
- **Calculating avoided emissions for regulatory purposes or a primary program objective.** Rigorous analyses are appropriate when avoided emissions are a primary goal of an efficiency program—typically, when the efficiency program is part of a regulatory scheme or is intended to generate creditable emission reductions or offsets with a significant monetary value. In this situation, documentation should be provided (either on a project-by-project basis or, preferably, on a program level) that the energy savings and avoided emissions are additional. A boundary assessment is also desirable to document that there is no "leakage," although in the case of most efficiency programs the boundary definition is straightforward. The energy savings used in the analyses should be net savings, with the net savings calculated to include only those energy savings that are additional. In the case of regulatory mandated programs, the mechanism for calculating avoided emissions will probably be defined. In other situations the more rigorous methods described in Chapter 6 for calculating avoided emissions should be used. In any event, the uncertainty issues discussed in Section 3.6 need to be addressed for the avoided emissions calculations as well as the energy savings calculations.

The following documents provide some guidance with respect to greenhouse gas programs. Each is a product of the World Business Council for Sustainable Development (WBCSD) and/or the World Resources Institute (WRI) and is available at <http://www.wri.org/climate/>.

- *Guidelines for Quantifying GHG Reductions from Grid-Connected Electricity Projects*, published in August 2007.
- *GHG Protocol Corporate Accounting and Reporting Standard* (Corporate Standard), revised edition, published in March 2004.
- *GHG Protocol for Project Accounting* (Project Protocol), published in December 2005.

Examples of energy efficiency projects implemented for their greenhouse gas emission benefits can be found at the Climate Trust Web site: <http://www.climatetrust.org/>.

3.10 Notes

1. These considerations, especially "free ridership," are sometimes subsumed under the more comprehensive term "attribution."

2. As is discussed in Chapter 5, calculating net savings can be problematic because (a) aspects of the net savings evaluation process are inherently subjective and (b) it is difficult to credit one particular efficiency program with benefits when there are many influences on energy consumer behavior.

3. In theory, demand rates of consumption can be of interest for fuel (e.g., natural gas) savings measures, as well. In practice they are not a concern. This discussion of demand savings is limited to electrical demand. However, it is important to understand that demand savings at the end-user level do not necessarily translate into capacity savings at the transmission or generation level.

4. DEER can be accessed at <www.energy.ca.gov/deer/index.html>.

5. DR's relationship with energy efficiency and environmental impacts is discussed in "The Green Effect, How Demand Response Programs Contribute to Energy Efficiency and Environmental Quality," *Public Utilities Fortnightly*, March 2007, <http://www.fortnightly.com>. The National Action Plan for Energy Efficiency plans to release additional guidance on the coordination of energy efficiency and DR programs in 2008.

6. Market progression is when the rate of naturally occurring investment in efficiency increases and can be considered to erode the persistence of earlier first year savings. An example of a cause of market progression is energy price effects—higher energy costs resulting in higher levels of efficiency.

4: Calculating Gross Energy and Demand Savings

Chapter 4 begins by defining key terms and introducing the fundamentals of calculating gross energy and demand savings. The next section provides a more detailed description of each of the three options for calculating gross energy savings, including M&V, deemed savings, and large-scale data analysis. The final section describes the primary considerations for selecting a gross savings approach.

4.1 Basics of Calculating Gross Savings

There is no direct way of measuring gross energy or demand savings, since one cannot measure the absence of energy use. However, the absence of energy use (i.e., gross energy and demand savings) can be estimated by comparing energy use (and demand) before and after implementation of a program. Thus, the following equation applies for energy savings and demand:

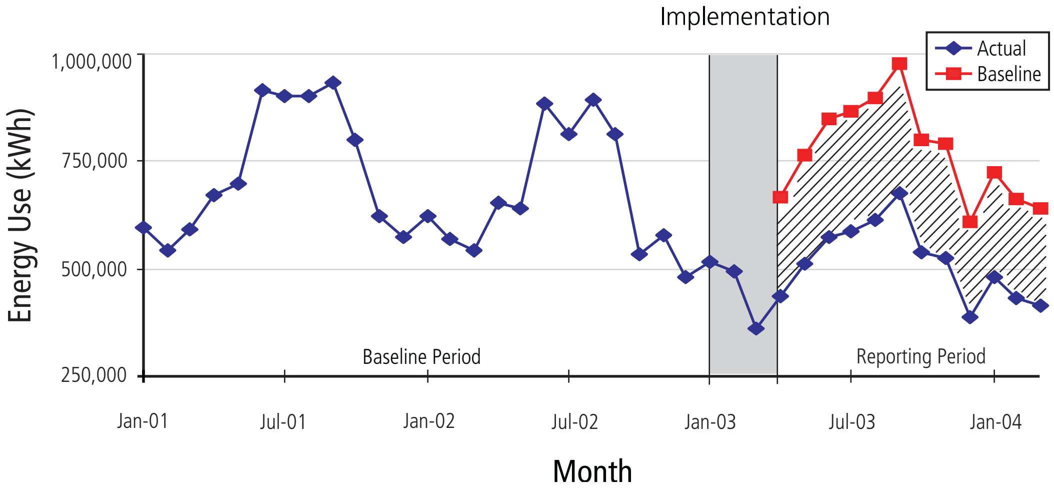
energy savings = (baseline energy use) – (reporting period energy use) ± (adjustments)

Weather Adjustments

The most common adjustment for comparing baseline and reporting period energy use in buildings is weather. This is because weather is often the primary independent variable for energy use in buildings. It is typically described in terms of ambient dry bulb temperature, the outdoor air temperature most people are familiar with seeing reported. It is reported in and described in terms of ˚F, cooling degree days (CDD), or heating degree days (HDD). CDD and HDD are common indicators of how space cooling or heating is required in a building, as a function of standard thermostat set points and outdoor air temperature. Other weather parameters that can be important include solar insolation and wet bulb temperature, which is an indication of ambient air temperature and humidity. Data on weather, both real-time and historical, are available from private companies and the National Oceanic and Atmospheric Administration (NOAA). See the IPMVP and ASHRAE Guideline 14 for more information on weather adjustments.

- **Baseline energy use** is the energy consumption estimated to have occurred before the program was implemented and is chosen as representative of normal operations. It is sometimes referred to as "*business-as-usual*" (BAU) energy use. When discussed in terms of specific projects, it is sometimes called the *pre-installation* energy use.
- **Reporting period energy use** is the energy consumption that occurs after the program is implemented. When discussed in terms of specific projects, it is sometimes called the *post-installation* energy use.
- **Adjustments** distinguish properly determined savings from a simple comparison of energy usage before and after implementation of a program. By accounting for factors (independent variables) that are beyond the control of the program implementer or energy consumer, the adjustments term brings energy use in the two time periods to the same set of conditions. Common examples of adjustment are:
  - Weather corrections—for example, if the program involves heating or air-conditioning systems in buildings.
  - Occupancy levels and hours—for example, if the program involves lighting retrofits in hotels or office buildings.
  - Production levels—for example, if the program involves energy efficiency improvements in factories.

Figure 4-1. Comparison of Energy Use Before and After a Program Is Implemented



The basic approach to evaluation is shown in Figure 4-1. It involves projecting energy use patterns of the baseline period into the reporting period. Such a projection requires adjustment of baseline energy use to reporting period conditions (weather, production level, occupancy, etc.). Therefore, the evaluation effort will involve defining the baseline energy use, the reporting period energy use, and any adjustments made to the baseline energy use.

A major impact evaluation decision is selecting the baseline. The baseline defines the conditions, including energy consumption and related emissions, that would have occurred without the subject program. The selection of a baseline scenario always involves uncertainty because it represents a hypothetical scenario.

Similarly, avoided emissions are calculated as those that result from a project or program that are additional to any that would have occurred in the absence of the project or program activity. This concept of "additionality" and the concepts of baselines used for calculating energy and demand savings are closely linked. While it is possible to have one baseline for calculating energy and demand savings and another for calculating avoided emissions, it is preferable to define a single baseline.

Baseline definitions consist of (a) site-specific issues and (b) broader, policy-orientated considerations. For each of these options, the two generic approaches to defining baselines are the project-specific and the performance standard procedure. These options and considerations for selecting one or the other, as well as considerations for selecting baseline adjustment factors, are discussed in the planning chapter (Section 7.2.4).

4.2 Measurement and Verification Approach

M&V is the process of using measurements to reliably determine actual savings created within an individual facility. This includes data collection as well as monitoring and analysis associated with the calculation of gross energy and demand savings. M&V covers all field activities dedicated to collecting site information, including equipment counts, observations of field conditions, building occupant or operator interviews, measurements of parameters, and metering and monitoring.

M&V involves determining gross energy and/or demand savings by:

- Selecting a representative sample of projects.
- Determining the savings of each project in the sample, using one or more of the M&V Options defined in the IPMVP.
- Applying the sample projects' savings to the entire population (i.e., the program).

Field Inspections of Energy Efficiency Measures

Not all of the evaluation approaches described in this chapter require field inspections, but typically there are some physical assessments for at least a sample of the individual projects in a program (i.e., field activities). This is to ensure that the measures installed meet appropriate specifications and that the projects included in a program have the potential to generate savings. This potential to generate savings can be verified through observation, inspections, and spot or short-term metering conducted immediately before and after installation. These field activities can also be conducted at regular intervals during the reporting period to verify a project's continued potential to generate savings. The field activities are an inherent part of the data collection aspects of the M&V approach, though they may be considered "add-ons" to the other approaches.

In the impact evaluation planning process, the M&V Option selected and some M&V details will need to be specified. In addition, each project evaluated will need to have a project-specific M&V plan. There are two types of project-specific M&V plans:

- **Prescriptive method plans**—for projects with significant M&V "experience" and well-understood determinants of savings (e.g., lighting and motor retrofits) there are established M&V procedures, example plans, and spreadsheets. The FEMP Guidelines contain prescription approaches to several common energy efficiency measures. ASHRAE Guideline 14 contains a prescriptive method for Option C, whole-facility analysis.[1]
- **Generic method plans**—conceptual approaches applicable to a variety of project types for which deemed values cannot be established and for which prescriptive M&V methods are not available (e.g., comprehensive building retrofits and industrial energy efficiency measures). The FEMP and ASHRAE Guidelines contain several generic methods and the 2007 IPMVP defines four generic methods, called Options.

The four IPMVP Options provide a flexible set of methods (Options A, B, C, and D) for evaluating energy savings in facilities. Having four options provides a range of approaches to determine energy savings with varying levels of savings certainty and cost. A particular Option is chosen based on the specific features of each project, including:

- Type and complexity.
- Uncertainty of the project savings.
- Potential for changes in key factors between the baseline and reporting period.
- Value of project savings.

This is because the Options differ in their approach to the level, duration, and type of baseline and reporting period measurements. For example, in terms of measurement levels:

- M&V evaluations using Options A and B are made at the end-use, system level (e.g., lighting, HVAC).
- Option C evaluations are conducted at the whole-building or whole-facility level.
- Option D evaluations, which involve computer simulation modeling, are also made at the system or the whole-building level.

In terms of type of measurement:

- Option A involves using a combination of both stipulations and measurements of the key factors needed to calculate savings in engineering models.
- Options B and C involve using spot, short-term, or continuous measurements[2] in engineering models (Option B) or regression analyses (Option C).
- Option D may include spot, short-term, or continuous measurements to calibrate computer simulation models.

Table 4-1. IPMVP Options (as Indicated in the 2007 IPMVP)

M&V Option	How Savings Are Calculated	Cost per Project (Not from IPMVP)	Typical Applications
A. Retrofit Isolation: Key Parameter Measurement Savings are determined by field measurement of the key performance parameter(s) that define the energy use of the efficiency measures' affected system(s) and the success of the project. Measurement frequency ranges from short-term to continuous, depending on the expected variations in the measured parameter and the length of the reporting period. Parameters not selected for field measurement are estimated. Estimates can based on historical data, manufacturer's specifications, or engineering judgment. Documentation of the source or justification of the estimated parameter is required. The plausible savings error arising from estimation rather than measurement is evaluated.	Engineering models of baseline and reporting period energy from short-term or continuous measurements of key operating parameter(s). Estimated values also used. Routine and non-routine adjustments as required.	Dependent on number of measurement points. Approximately 1% to 5% of project construction cost of items subject to M&V.	A lighting retrofit where power draw is the key performance parameter that is measured. Estimate operating hours of the lights based on building schedules, occupant behavior, and/or prior studies.
B. Retrofit Isolation: All Parameter Measurement Savings are determined by field measurement of the energy use of the affected system. Measurement frequency ranges from short-term to continuous, depending on the expected variations in the savings and the length of the reporting period.	Short-term or continuous measurements of baseline and reporting-period energy, and/or engineering models using measurements of proxies of energy use. Routine and non-routine adjustments as required.	Dependent on number and type of systems measured and the term of analysis or metering. Typically 3% to 10% of project construction cost of items subject to M&V.	Application of a variable-speed drive and controls to a motor to adjust pump flow. Measure electric power with a meter installed on the electrical supply to the motor, which reads the power every minute. In the baseline period this meter is in place for a week to verify constant loading. The meter is in place throughout the reporting period to track variations in power use.

Source: EVO, 2007.

Table 4-1 (continued). IPMVP Options (as Indicated in the 2007 IPMVP)

M&V Option	How Savings Are Calculated	Cost per Project (Not from IPMVP)	Typical Applications
C. Whole Facility Savings are determined by measuring energy use at the whole-facility or sub-facility level. Continuous measurements of the entire facility's energy use are taken throughout the reporting period.	Analysis of whole-facility baseline and reporting period (utility) meter data. Routine adjustments as required, using techniques such as simple comparison or regression analysis. Non-routine adjustments as required.	Dependent on number and complexity of parameters in analysis and number of meters. Typically 1% to 5% of project construction cost of items subject to M&V.	Multifaceted energy management program affecting many systems in a facility. Measure energy use with the gas and electric utility meters for a 12-month baseline period and throughout the reporting period.
D. Calibrated Simulation Savings are determined through simulation of the energy use of the whole facility, or of a sub-facility. Simulation routines are demonstrated to adequately model actual energy performance measured in the facility.	Energy use simulation, calibrated with hourly or monthly utility billing data. (Energy end-use metering may be used to help refine input data.)	Dependent on number and complexity of systems evaluated. Typically 3% to 10% of project construction cost of items subject to M&V.	Multifaceted, new construction, energy management program affecting many systems in a facility—applies where no meter existed in the baseline period. Energy use measurements, after installation of gas and electric meters, are used to calibrate a simulation. Baseline energy use, determined using the calibrated simulation, is compared to a simulation of reporting period energy use.

Source: EVO, 2007.

The four generic M&V options are summarized in Table 4-1. While these options are directly associated with energy efficiency projects, the basic concepts are also applicable to water conservation, clean power, transportation, and distributed generation activities.

One of the key aspects of M&V is defining a *measurement boundary*. The measurement boundary might be a single piece of equipment (e.g., the replaced motor in a factory), a system (e.g., the entire lighting system retrofitted in a commercial building), or the whole facility (e.g., for a home that has undergone a complete retrofit).

Any energy effects occurring beyond the measurement boundary are called "interactive effects." A typical interactive effect is the decrease in air-conditioning requirements or increase in space heating requirements that can result from a lighting retrofit, which by its nature reduces the amount of heat produced by a lighting system. The magnitude of such interactive effects, if significant, should be considered and a measurement method developed to estimate them under the savings determination process.

4.2.1 M&V Option A: Retrofit Isolation—Key Parameter Measurement

Option A involves project- or system-level M&V assessments in which the savings associated with a particular project can be isolated. With this Option, key performance parameters or operational parameters can be spot or short-term measured during the baseline and reporting periods. However, some parameters are stipulated rather than measured. This level of verification may suffice for types of projects in which a single parameter represents a significant portion of the savings uncertainty.

Under Option A, energy and demand savings are calculated using "engineering models." These models are essentially groups of equations defining energy use as a function of various inputs—often simple spreadsheet models—and involve developing estimates of energy and demand savings based on:

- Assumptions concerning operating characteristics of the equipment or facilities in which the equipment is installed, which are informed by measurements (from spot to continuous). Examples are power draws (wattage) of light fixture or fan motors and efficiencies of air-conditioners (kWh/ton) and heaters (Btu out/Btu in).
- Assumptions about how often the equipment is operated or what load it serves. Examples are operating hours of lights or fixed-speed fans and air conditioning loads (tons) or heater loads (Btu).

The most straightforward application of engineering models involves using savings algorithms that summarize how energy use is expected to change due to installation of the energy efficiency measure. Savings are then estimated by changing the model parameters that are affected by program participation. With Option A, at least one of the key model parameters must be measured. The parameters not measured are stipulated based on assumptions or analysis of historical or manufacturer's data. Using a stipulated factor is appropriate only if supporting data demonstrate that its value is not subject to fluctuation over the term of analysis.

Interactive Factors

Interactive effects are those that an energy efficiency measure has on energy use in a facility, but which are indirectly associated with the measure. For example, reduction in lighting loads through an energy-efficient lighting retrofit will reduce air conditioning and/or increase heating requirements, since there is less heat generated by the energy-efficient lights. When energy efficiency programs have interactive effect beyond a single building and start to impact energy supply and distribution systems, there can be implications for calculating avoided emissions and other related co-benefits. In this situation of wide-scale interactive effects, the term "leakage" is used.

This Option, and Option B, are best applied to programs that involve retrofitting equipment or replacing failed equipment with efficient models. All end-use technologies can be verified using Option A or B; however, the validity of this Option is considered inversely proportional to the complexity of the measure. Thus, the savings from a simple lighting retrofit (less complex) may be more accurately determined with Option A or B than the savings from a chiller retrofit (more complex).

Also true with Options A and B is that measurement of all end-use equipment or systems may not be required if statistically valid sampling is used. For example, the operating hours for a selected group of lighting fixtures and the power draw from a subset of representative constant-load motors may be metered.

Savings determinations under Option A can be less costly than under other Options, since the cost of deriving a stipulation is usually less than the cost of making measurements. However, since some stipulation is allowed under this Option, care is needed to review the engineering design and installation to ensure that the stipulations are realistic and achievable (i.e., the equipment can truly perform as assumed). At defined intervals during the reporting period, the installation can be re-inspected to verify the equipment's continued existence and its proper operation and maintenance. Such re-inspections will ensure continuation of the potential to generate predicted savings and validate stipulations.

4.2.2 M&V Option B: Retrofit Isolation—All Parameter Measurement

Option B, like Option A, involves project- or system-level M&V assessments with performance and operational parameters measured at the component or system level. Option B also involves procedures for verifying the potential to generate savings that are the same as Option A. In addition, savings calculations, as with Option A, involve the use of engineering models. *However, unlike Option A, Option B does not allow stipulations of major factors*.

Thus, Option B requires additional and often longer-term measurements compared to Option A. These include measurements of both equipment operating characteristics (as would be required under Option A) and relevant performance factors. Commonly measured parameters include operating hours for lighting and HVAC equipment, wattage for lighting and HVAC equipment, and line flows and pressure for various compressed air applications.

Option B relies on the direct measurement of end-uses affected by the project. Spot or short-term measurements may be sufficient to characterize the baseline condition. Short-term or continuous measurements of one or more parameters take place after project installation to determine energy use during the reporting period.

All end-use technologies can be verified with Option B, but the difficulty and cost increase as measurement complexity increases. Measuring or determining energy savings using Option B can be more difficult than doing so with Option A. The results, however, are typically more reliable. In addition, the use of longer-term measurements can help identify under-performing efficiency projects, which in turn can lead to improvements in their performance.

Retrofit Isolation and Measurement Boundaries Example

A factory's boiler, used for process steam production, is replaced with a more efficient boiler of about the same capacity. The *measurement boundary* is defined to just include the boiler, whether the baseline boiler (before it is replaced) or the more efficient boiler (once it is installed). With this boundary, the analyses of baseline and efficient boilers are not affected by variations in the factory's process steam load, although the actual savings depend on the steam consumption of the factory. Meters for fuel consumption and boiler steam output are all that are needed to assess the efficiencies of the baseline and efficient boilers over their full range of operations.

Under Option A, *savings* are reported for the boiler retrofit by applying the measured annual average efficiency improvement to an estimated annual boiler load and the boiler efficiency test is repeated annually during the reporting period.

Under Option B, the annual boiler load may be determined by measuring the boiler load over several weeks (to prepare typical hourly and daily load profiles) and then using this information to make a more accurate savings estimate based on matching typical hourly load profiles to partial and full steam load boiler efficiency profiles, rather than just using an annual average efficiency value and an average annual steam consumption value.

4.2.3 M&V Option C—Whole-Facility Analyses

Option C involves use of whole-building meters or sub-meters to assess the energy performance of a total building or facility. These meters are typically the ones used for utility billing, although other meters, if properly calibrated, can also be used. Option C is the most common form of M&V for building energy efficiency retrofits. With this option, energy consumption from the baseline period is compared with energy consumption bills from the reporting period. Option C involves procedures for verifying the potential to generate savings that are the same as Option A.

Whole-building or facility level metered data are evaluated using techniques ranging from simple bill comparisons to multivariate regression analysis. Option C regression methods can be powerful tools for determining savings, while simple bill comparison methods are *strongly discouraged*. The latter approach does not account for independent variables, such as weather.

For the regression analyses to be accurate, all explanatory (independent) variables that affect energy consumption need to be monitored during the performance period. Critical variables may include weather, occupancy schedules, throughput, control set points, and operating schedules. Most applications of Option C require at least 9 to 12 months of continuous baseline (pre-installation) meter data and at least 9 to 12 months of continuous data from the reporting period (post-installation).

For programs targeting integrated whole-building approaches to energy efficiency, utility bill analysis can be used to statistically evaluate persistence. One useful tool that can be used for this purpose is EPA's ENERGY STAR Portfolio Manager.

All end-use technologies can be verified with Option C. However, this option is intended for projects where savings are expected to be large enough to be discernible from the random or unexplained energy variations normally found at the level of the whole-facility meter. The larger the savings, or the smaller the unexplained variations in the baseline consumption, the easier it will be to identify savings. In addition, the longer the period of savings analysis after project installation, the less significant the impact of short-term unexplained variations. Typically, savings should be more than 10% of the baseline energy use so that they can be separated from the "noise" in baseline data.

EPA's Portfolio Manager

One tool that can be used to analyze facility utility billing meter data is EPA's Portfolio Manager (PM).[3] Over 30,000 buildings have been benchmarked with PM, which provides a consistent framework and metric that building energy managers can use to track, measure, and monitor whole-building energy use. PM employs a methodology that is consistent with IPMVP Option C. PM aggregates all the meter data from a building so that performance changes can be assessed at the whole-facility level. Savings are determined at the building level to promote system-wide energy reductions. Additionally, because the PM approach combines multiple meters it accounts for differences among fuel types. This is done by converting site meter data into source energy (or, "primary energy") consumption.

4.2.4 M&V Option D—Calibrated Simulation

Option D involves calibrated computer simulation models of systems, system components, or whole-facility energy consumption to determine project energy savings. Linking simulation inputs and results to baseline or reporting period data calibrates the results to actual billing or metered data. Typically, reporting period energy use data are compared with the baseline computer simulation energy use prediction (using reporting period independent variable values) to determine energy savings.

Manufacturer's data, spot measurements, or short-term measurements may be collected to characterize baseline and reporting period conditions and operating schedules. The collected data serve to link the simulation inputs to actual operating conditions. The model calibration is accomplished by comparing simulation results with end-use or whole-building data. Whole-building models usually require at least 9 to 12 months of pre-installation data for baseline model calibration. However, these models are sometimes calibrated with only reporting period data so that they can be used with new construction projects for which no baseline data exist.

Building Energy Simulation Programs

For over 30 years, engineers and scientists have been developing computerized models that describe how the energy use of buildings changes in response to independent variables, such as weather. The sophistication and complexity of these models is quite varied. To learn about some of the building simulation models that are publicly available, see the Lawrence Berkeley National Laboratory Simulation Research Group Web page at <http://gundog.lbl.gov/> and the Texas Energy Systems Laboratory Web page at <http://esl.eslwin.tamu.edu/>.

Any end-use technology can be verified with Option D if the drop in consumption is larger than the associated simulation modeling error. This option can be used in cases where there is a high degree of interaction among installed energy systems, or where the measurement of individual component savings is difficult. Option D is commonly used with new construction energy efficiency programs, where the baseline is typically modeled using standard practice or building code requirements to define what would have occurred without the efficiency activity.

Savings determined with Option D are based on one or more complex estimates of energy use. Therefore, the quality of the savings estimate depends on how well the simulation models are calibrated and how well they reflect actual performance. Since building simulation models can involve elaborate spreadsheets or vendor estimating programs, accurate modeling and calibration are the major challenges associated with Option D.

4.3 Deemed Savings Approach

Deemed savings are used to stipulate savings values for projects with well-known and documented savings values. Examples are energy-efficient appliances such as washing machines, computer equipment and refrigerators, and lighting retrofit projects with well-understood operating hours. Several programs use stipulated values, as well as other mechanisms, for determining individual project and thus program savings. These include the NYSERDA (New York) Energy $mart Program, the Texas DSM programs, and the California standard offer programs, which have prepared deemed savings values for certain measure types. For these programs, deemed savings are used for only pre-qualified measures.[4]

The use of deemed values in a savings calculation is essentially an agreement between the parties to an evaluation to accept a *stipulated value*, or a set of assumptions, for use in determining the baseline or reporting period energy consumption. With the deemed savings approach, it is increasingly common to hold the stipulated value constant regardless of what the actual value is during the term of the evaluation. If certain requirements are met (e.g., verification of installation, satisfactory commissioning results, annual verification of equipment performance, and sufficient equipment or system maintenance), the project savings are considered to be confirmed. The stipulated savings for each verified installed project are then summed to generate a program savings value. Installation might be verified by physical inspection of a sample of projects or perhaps just an audit of receipts.

Deemed values, if used, should be based on reliable, traceable, and documented sources of information, such as:

- Standard tables from recognized sources that indicate the power consumption (wattage) of certain pieces of equipment that are being replaced or installed as part of a project (e.g., lighting fixture wattage tables).
- Manufacturer's specifications.
- Building occupancy schedules.
- Maintenance logs.

When using deemed values, it is important to realize that technologies alone do not save energy; it is how they are used that saves energy. Therefore, a deemed energy savings value depends on how and where a technology is placed into use. For example, a low-wattage lamp's savings are totally dependent on its operating hours. Such a lamp installed in a closet will save much less energy than one installed in a kitchen.

The example of the residential lamp raises the issue of "granularity" of the deemed savings values. In that example, if an average household's annual operating hours were used, the result would be underestimated savings if lamps were only installed in high-use areas and overestimated savings if lamps were only installed in low-use areas. Thus, the value of deemed savings depends not only on the validity of the value used, but on whether the value is applied correctly—that is, it must be based on the use conditions as well as the technology.

Sources of stipulated values must be documented in the evaluation plan. Even when stipulated values are used in place of measurements, equipment installation and proper operation are still verified. Properly used, stipulations can reduce M&V costs and simplify procedures. Improperly used, they can give evaluation results an inappropriate aura of authority. Deciding whether parameters could be stipulated requires understanding how they will affect savings, judging their effect on the uncertainty of results, and balancing the costs, risks, and goals of the program being evaluated.

Assessing a few key aspects of the project could drive decisions about whether to use stipulations and how to use them effectively in an evaluation plan:

- Availability of reliable information.
- The project's likelihood of success in achieving savings.
- Uncertainty of the stipulated parameter and its contribution to overall project uncertainty.
- The cost of measurement.

Uncertainty in predicted savings, and the degree to which individual parameters contribute to overall uncertainty, should be carefully considered in deciding whether to use stipulations. Savings uncertainty can be assessed by identifying the factors that affect savings and estimating the potential influence of each factor. Factors having the greatest influence should be measured if at all practical. Several "rules of thumb" are:

- The most certain, predictable parameters can be estimated and stipulated without significantly reducing the quality of the evaluation results.
- Stipulating parameters that represent a small degree of uncertainty in the predicted result and a small amount of savings will not produce significant uncertainty concerns.
- Parameters could be measured when savings and prediction uncertainty are both large.
- Even if savings are high, but uncertainty of predicted savings is low, full measurement may not be necessary for M&V purposes.

4.4 Large-Scale Data Analysis Approach

Large-scale data analysis applies a variety of statistical methods to measured facility energy consumption meter data (almost always whole-facility utility meter billing data) and independent variable data to estimate gross energy and demand impacts.[5] Unlike the M&V whole-facility analysis option (IPMVP Option C) described in Section 4.2, the meter analysis approach usually (a) involves analysis of a census of project sites, versus a sample, and (b) does not involve onsite data collection for model calibration—although inspections of a sample of projects to confirm proper operation of installed measures are still performed.

Most analyses of meter data involve the use of comparison groups (which can be hard to find in areas with a long history of program offerings). In assessing the impacts of programs, evaluators have traditionally used "quasi-experimental design." They compare the behavior of the participants to that of a similar group of nonparticipants—the comparison group—to estimate what would have happened in the absence of the program. The two groups need to be similar on average. The only difference should be the fact that one participated in an energy efficiency program and one did not. The observed change in consumption in the comparison group can be assumed to resemble the change in consumption that would have been observed in the participant group had it not been through a program.

There are three basic large-scale meter data analysis methods employed for energy efficiency programs:

- **Time series comparison**—compares the program participants' energy use before and after their projects are installed. With this method the "comparison group" is the participants' pre-project consumption. Thus, this method has the advantage of not requiring a comparison group of non-participants. The disadvantages are that it cannot be easily applied to new construction programs and even with well-established regression techniques, this approach cannot fully account for all changes in all the independent variables that might impact energy savings. The basic evaluation equation is:

$$savings = Q_{pre\text{-}installation} - Q_{post\text{-}installation}$$

where: $Q_{pre\text{-}installation}$ = quantity of energy used before the projects were implemented, corrected for independent variables, such as weather, to match reporting period independent variable values

$Q_{post\text{-}installation}$ = quantity of energy used after the projects were implemented

- **Use of comparison group**—compares the program participants' energy use after projects are installed with the energy use of non-participants. This method is used primarily for new construction programs, where there are no baseline data. The difficulty with this approach is usually related to the cost of analyzing two groups and finding a comparison group with sufficiently similar characteristics to the group of participants. The basic evaluation equation is:

$$savings = Q_{non\text{-}participants} - Q_{participants}$$

where: $Q_{participants}$ = quantity of energy used by the participants after their projects are installed

$Q_{non\text{-}participants}$ = quantity of energy used by the control group of non-participants, after the participants installed their projects

- **Comparison group/time-series**—this approach combines the two above approaches and thus has the advantages of comparing similar if not identical groups to each other while accounting for efficiency savings that would have occurred irrespective of the program. If the participant and comparison group are available, it is a preferred approach. The basic evaluation equation is:

$$savings = (Q_{pre\text{-}installation} - Q_{post\text{-}installation})_{participants} - (Q_{pre\text{-}installation} - Q_{post\text{-}installation})_{non\text{-}participants}$$

where: $Q_{pre\text{-}installation}$ = quantity of energy used before the projects were implemented

$Q_{post\text{-}installation}$ = quantity of energy used after the projects were implemented

Statistical models apply one of a number of regression analysis techniques to measured energy use data to control for variations in independent variables. With regression analyses, a relationship is defined (in the form of an equation or group of equations) between the dependent variable and one or more important independent variables. Dependent variables are the output of an analysis. Independent variables are the variables which are presumed to affect or determine the dependent variables and are thus inputs to an analysis. In the case of energy efficiency analyses, the output is energy or demand consumption and savings. The analysis itself is done with a computer model, which can be anything from a spreadsheet tool to sophisticated proprietary statistical modeling software.

The primary consideration for any evaluation is that the analysis must be designed to obtain reliable energy savings. Uncertainty of savings estimates can decrease as the evaluators attempt to incorporate the major independent variables that may have affected the observed change in consumption. This can be accomplished in several ways. One common method is to include participant and non-participant analyses (the second and third bullets above). If one of these approaches 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, evaluation 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.

It is very important to note that simple comparison of meter data—say subtracting this year's utility bills from the utility bills from before the measure installations—is not a valid evaluation approach (equation 4.1 above shows that the baseline data are corrected for the changes in independent variables). Simple comparison of reporting period energy use with baseline energy use does not differentiate between the effects of a program and the effects of other factors, such as weather. For example, a more efficient air conditioner may consume more electricity after its installation if the weather is warmer during the reporting period than it was before installation. To isolate the effects of the evaluated program (i.e., to establish attribution), the influence of these complicating factors must be addressed through the use of regression analyses.

In regression analysis, the following questions need to be answered:

- What independent variables are relevant to calculating energy savings? Often this is decided by common sense, experience, or budget considerations (with respect to how many variables can be measured and tracked) but it can also be determined through field experiments and statistical tests. For weather data (the most common independent variable), there is a wide range of public and private data sources.

- Will a comparison group be used in the analysis? While often a more accurate approach, the use of comparison groups assumes that a comparable group of participants and non-participants can be found and analyzed. This, of course, adds to evaluation costs.

- How will the analysis be tested for statistical errors, and what level of uncertainty is acceptable? The first concern requires qualified analysts and a quality control system. The second requires specification of statistical parameters that define the uncertainty of the calculated savings. The field of statistical analysis can be quite complex and untrained analysts often misinterpret analyses and miss key considerations or errors in statistical analyses.

- Are gross or net savings values desired? The latter two methods described above, which include comparison groups, can actually produce net savings values.

In addition, the appropriate type of statistical model needs to be decided. The following are brief descriptions of some typical generic model types:

- **Normalized annual consumption (NAC) analysis.** This is a regression-based method that analyzes monthly energy consumption data. 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.[6] The NAC method, often using PRISM, has been most often used to estimate energy impacts produced by whole-house retrofit programs.

- **Conditional savings analysis (CSA).** CSA is a type of analysis in which change in consumption is modeled using regression analysis against the presence or absence of energy efficiency measures. These are usually entered in the form of binary variables (1 if measures are installed and 0 if not).

- **Statistically adjusted engineering (SAE) models.** A category of statistical analysis models that incorporate the engineering estimate of savings as a dependent variable. For example, a SAE model can use change in energy as the dependant variable in a regression model against estimated savings for installed efficiency measures. Often these estimates are provided in the design phase or through secondary sources (e.g., DEER). When the measures are installed, the estimated savings is entered as the explanatory variable value. When the measures are not installed, 0 is entered as the explanatory variable value in the regression model.

- **Analysis of covariance (ANCOVA) models.** These are also called fixed effects models. Any of the above can be run as an ANCOVA model. The advantage of this approach is that it allows each participant or non-participant to have a separate estimate of the "intercept" term. Regression models estimate an intercept (in the case of energy modeling, this often represents the base component, i.e., non-weather sensitive component of energy use) and a slope coefficient (this often represents the change in energy consumption for one unit change in the explanatory variable). By permitting each participant and non-participant to have its own intercept, analysts allow for some differences among the analysis subjects.

While this Guide does not delve into statistical modeling details, an excellent source of information on the techniques described below is *The 2004 California Evaluation Framework* (CPUC, 2004).

4.5 Selecting a Gross Savings Evaluation Approach

Selecting an evaluation approach is tied to objectives of the program being evaluated, the scale of the program, evaluation budget and resources, and specific aspects of the measures and participants in the program. The following subsections describe situations in which each of the three gross impact approaches is or is not applicable.

One criterion that works across all of the approaches is evaluator experience and expertise. Thus, a common requirement is that the evaluator has experience with the approach selected.

4.5.1 M&V Approach

The M&V approach is used for almost any type of program that involves retrofits or new construction projects. While a census of projects can be used with the M&V approach, it is generally applied to only a sample of projects in a program. This is because the M&V approach tends to be more expensive on a per-project basis than the other two approaches. In general, the M&V approach is applied when the other approaches are not applicable or when per-project results are needed. An example is a performance-contracting program with multiple contractors.

Because the selection of the M&V approach is contingent on which of the four M&V Options is selected, Table 4-2 summarizes some selection criteria for each M&V Option. Cost is one of these considerations and is influenced by the following factors:

Option A

- Number of measurement points
- Complexity of deriving the stipulation
- Frequency of post-retrofit inspections

Option B

- Number of points and independent variables measured
- Complexity of measurement systems
- Length of time measurement system maintained
- Frequency of post-retrofit inspections

Option C

- Number of meters to be analyzed
- Number of independent variables used in models

Option D

- Number and complexity of systems simulated
- Number of field measurements required for model input data

- Effort required for calibration of model

Figure 4-2 is taken directly from the 2007 IPMVP and shows a flowchart summarizing the selection of M&V Options. It can be used to select the IPMVP Option, or Options, that are appropriate for a given program.

4.5.2 Deemed Savings Approach

Deemed savings approaches are most commonly used for programs that involve simple new construction or retrofit energy efficiency measures with well-defined applications. Examples might be T-8 lighting retrofits in office buildings or compact CFL giveaways for residential utility customers. In each of these two examples, an assumption would be made about the average wattage savings and the average hours of operation combined with the effort of verifying that the T-8s were installed and the CFLs actually provided to residents. Deemed values would be based on historical evaluations of other similar programs.

In general, the deemed savings approach is most applicable when all or at least most of the following are true:

- There are limited evaluation resources.
- The projects involve simple energy efficiency measures with well-understood savings mechanisms, and are not subject to significant variation in savings due to changes in independent variables.
- The uncertainty associated with savings estimates is low and/or the risk of under- (or over-) estimating savings is low.
- Documented per-measure stipulated values are available and applicable to the measure installation circumstances.
- The primary goal of the evaluation is to conduct field inspections for all or a sample of projects, to make sure they are properly installed and have the potential to generate savings (rather than having rigorously determined energy savings).

4.5.3 Large-Scale Data Analysis Approach

These approaches are most commonly used for programs that involve large-scale retrofit programs with many participants. A typical example is a residential customer weatherization program with thousands of homes being retrofitted with a variety of measures such as insulation, weather stripping, low-flow showerheads, and compact fluorescent light bulbs (CFLs). In general, the large-scale data analysis approach is most applicable to programs that meet most if not all of the following criteria:

- Participation is well defined (i.e., the specific customers or facilities that participated in the program are known).
- The program has a relatively large number of participants (i.e., probably over 100).
- At least one year's worth of energy consumption data are available after program measures are installed. If a comparison group is not used, at least one year's worth of baseline energy consumption data should also be available. Depending on the quality of the available data, a shorter data period may be adequate (i.e., if daily, versus monthly, data are used).
- There is some similarity between participants, or relatively homogenous subgroups of participants can be formed with similar facility and energy efficiency measure characteristics.
- Expected changes in energy consumption due to measures installed through the program account for at least 10 percent of facility energy consumption (preferably more than 15 percent).

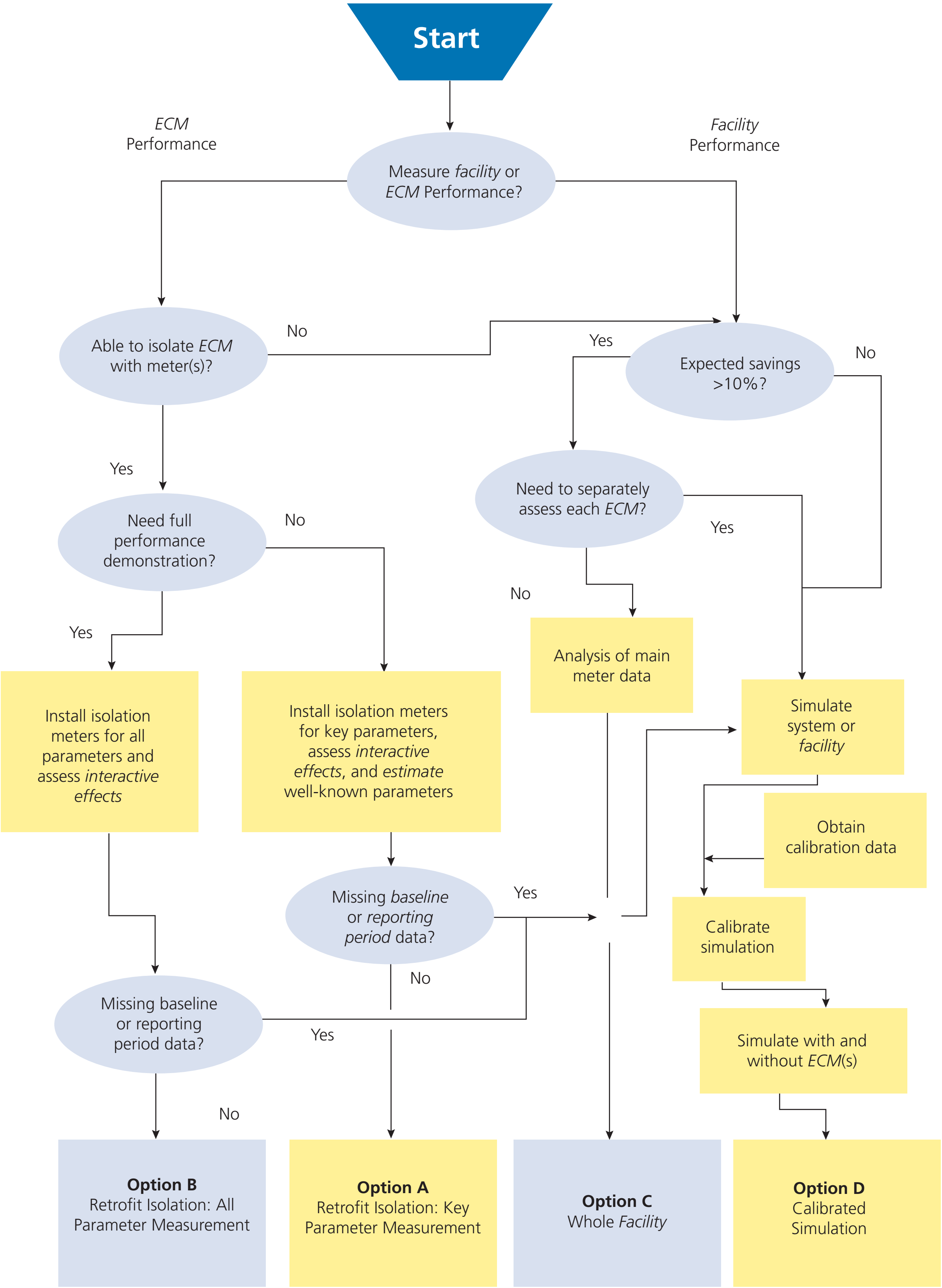
This approach can be used with both retrofit and new construction programs and is generally applied to a census of the projects in a program.

Table 4-2. Applications for Each IPMVP M&V Option

Option A Retrofit Isolation— Key Parameter Measurement is best applied where:	**Option B Retrofit Isolation— All Parameters Measurement** is best applied where:	**Option C Whole Facility** is best applied where:	**Option D Calibrated Simulation** is best applied where:
• The magnitude of savings is low for the entire project or for the portion of the project to which Option A is applied • The project is simple, with limited independent variables and unknowns • The risk of not achieving savings is low • Interactive effects are to be ignored or are stipulated using estimating methods	• The project involves simple equipment replacements • Energy savings values per individual measure are desired • Interactive effects are to be ignored or are stipulated using estimating methods • Independent variables are not complex	• The project is complex • Predicted savings are large (typically greater than 10%) compared to the recorded energy use • Energy savings values per individual measure are not needed • Interactive effects are to be included • Independent variables that affect energy use are not complex or excessively difficult to monitor	• New construction projects are involved • Energy savings values per measure are desired • Option C tools cannot cost-effectively evaluate particular measures • Complex baseline adjustments are anticipated • Baseline measurement data do not exist or are prohibitively expensive to collect

Source: EVO, 2007.

Figure 4-2. IPMVP M&V Option Selection



4.6 Notes

1. See Appendix E for information on ASHRAE Guideline 14, the FEMP M&V Guideline, and the IPMVP.
2. Spot measurements are one-time measurements, for example of the power draw of a motor. Short-time measurements might take place for a week or two, such as to determine the operating hours of lights in an office. Continuous measurements, as the name implies, involve measuring key factors such as power consumption or outdoor temperature throughout the term of the evaluation, which may be years.
3. See <http://www.energystar.gov/index.cfm?c=evaluate_performance.bus_portfoliomanager>.
4. NYSERDA: <http://www.nyserda.org>; Texas: <http://www.puc.state.tx.us/electric/projects/22241/22241arc/CI-MV.doc>; California: <http://eega.cpuc.ca.gov/deer/>.
5. As discussed in Chapter 5, related analyses can be used to also calculate net savings.
6. PRISM: <http://www.princeton.edu/~marean/>. SAS: <http://www.sas.com/>. SPSS: <http://www.spss.com/>.

5: Calculating Net Energy and Demand Savings

Chapter 5 defines net savings and describes the four key factors that differentiate net and gross savings: free ridership, spillover effects, rebound effects, and electricity transmission and distribution losses. The chapter then provides a detailed description of several approaches for determining net savings, including self-reporting surveys, econometric models, and stipulated net-to-gross ratios. A brief discussion of the criteria for selecting an appropriate net savings evaluation approach is also provided.

5.1 Importance of Net Savings

To keep program benefits from being under- or overstated, it is important to understand and properly reflect the influences of both energy savings and emission avoidance programs. These net savings are the savings "net" of what would have occurred in the absence of the program. Generally speaking, net savings are of most interest for regulated government and utility programs. In these cases, the responsible party (for example, a city council or utility regulator) wants to know if the use of public or ratepayer funded programs are actually having an influence. That is, are the programs of interest providing incremental benefits, or do the benefits result from some other influences? For example, the environmental benefits of energy efficiency programs are usually considered valid only if they are additional to naturally occurring efficiency activities (that is, based on net savings). In contrast, private sector energy efficiency programs such as performance contracts are a case where gross energy savings are the primary concern.

The following sections describe factors that differentiate net and gross impacts and approaches for calculating NTGRs. It is important to understand, though, that calculating net energy and demand savings can be more of an art than a science. Essentially, one is attempting to separate out the influence of a particular energy efficiency program (or portfolio) from all the other influences that determine participant and non-participant behavior and decisions. With the increasing "push" for energy efficiency by utilities and government at the local, state, and national level and by private groups and large companies, it can be quite difficult to separate out how one particular program among all this activity influences the decision of whether, when, and to what degree to adopt efficiency actions.

5.2 Factors That Account for Differences Between Net and Gross Savings

The three primary factors that differentiate gross and net savings are free ridership, spillover, and rebound. In addition, transmission and distribution losses can also be considered under a NTGR calculation for programs that save electricity from grid-connected power plants. The decision about which of these to include in an NTGR analysis is determined by the objectives of the evaluation. Free ridership is typically the most commonly evaluated NTGR factor, followed by spillover and then rebound analyses.

- **Free ridership.** Free riders are program participants who would have implemented the program measure or practice in the absence of the program. The program can also affect when a participant implements an efficiency measure (e.g., because of the program a participant installs the equipment sooner than he or she otherwise would have), the level of efficiency of the efficient equipment installed (e.g., a participant says he or she would have installed the same efficient equipment without the program), and the number of units of efficiency equipment installed. Different levels of free ridership introduce the concept of partial or deferred free riders. The subjectivity surrounding free ridership is a significant component of net energy and demand savings uncertainty.

Free Riders

There are three categories of free riders:

- **Total free rider**—would have installed the same energy efficiency measures at the same time whether or not the program existed.
- **Partial or deferred free rider**—would have installed less-efficient (but still more efficient than baseline) measures or would have installed the same energy efficiency measure but at a later time and would have installed fewer of the energy efficiency products.
- **Non-free rider**—would not have installed the baseline energy efficiency measure without the influence of the program.

It should be noted that a participant's free ridership status can vary from one measure to the next and over time.

- **Spillover effects.** Spillover occurs when there are reductions in energy consumption or demand caused by the presence of the energy efficiency program, but which the program does not directly influence. Customer behavioral changes stemming from participation in programs are a positive program spillover, increasing the program effect. 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 (and they purchase) as a result of program availability; (c) changes in specification practices employed by architects and engineers; and (d) 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). The term "free driver" is used to describe a non-participant who has adopted a particular efficiency measure or practice as a result of a utility program.

The analysis of spillover and free ridership is complicated by "market noise." When a market is filled with many implementers offering similar programs under different names, with different incentive structures and marketing methods, it is difficult to estimate any particular program's influence. Identification of non-participants may also be difficult, since customers may not be able to discern between the various programs operating in the marketplace and may not accurately recall how programs may have influenced their decision processes or even remember the program in which they participated.

- **Rebound effect.** Rebound is a change in energy-using behavior that increases the level of service and results from an energy efficiency action. The most common form is "take back," which can occur if consumers increase energy use as a result of a new device's improved efficiency. For example, homeowners may use more air-conditioning with their new efficient air-conditioner because it is cheaper to run than their old air-conditioner. Another example is when insulation is installed for a low-income household and the homeowner can turn the thermostat up to a more comfortable temperature. However, there is a non-energy benefit here associated with increased comfort, health, and safety that some would argue should be considered a co-benefit.

Rebound effect is part of the general concept of how customer behavior affects technology usage and, thus, efficiency performance. For example, installation of occupancy sensors in small independent hotels would not save energy if hotel staff were already adjusting HVAC manually as part of their ordinary maintenance. In another example, an Energy Management System could be overridden by management decisions. Behavioral issues such as these are becoming of increasing interest in advanced energy efficiency programs.

- **Electricity transmission and distribution losses.** When an efficiency project reduces electricity consumption at a facility, the amount of electricity that no longer has to be generated at a power plant is actually greater than the onsite reduction. This is because of electricity transmission and distribution (T&D) losses between the sites and the power plants. Published electricity grid emission factors do not usually include T&D losses and most energy savings evaluations only report onsite energy savings. Therefore an evaluator needs to decide whether to include T&D losses in their net savings calculation.

T&D losses can range from negligible for a high-voltage customer located close to a power plant to over 10% for smaller customers located far from power plants. In addition, higher T&D losses are inevitable during on-peak hours. Thus, some jurisdictions have calculated on-peak, off-peak, and seasonal T&D loss factors.

If a T&D loss factor is being considered, it is best to adopt one factor (or perhaps two, one for on-peak and one for off-peak) for the entire grid and not attempt to be too fine-grained. Two options for quantifying T&D losses are (a) assuming a simple percentage adder for source savings and (b) not including T&D losses directly, but considering them a counterweight to uncertainty in the site savings calculation. The adder could be a value calculated for the specific T&D network in question. Potential sources of such data are local regulatory authorities, local utilities, and the regional independent system operator (ISO).

EPA's Conservation Verification Protocol (EPA, 1995) for the Acid Rain Program suggests the following default values for T&D losses, as a proportional adder to onsite energy savings:

- T&D savings for residential and commercial customers—7 percent
- T&D savings for industrial customers—3.5 percent

This consideration of T&D issues is often part of a calculation to determine "source" energy (fuel) savings (i.e., how much fuel is not consumed in a power plant because of the end-use efficiency activity).

Source fuel savings are calculated by considering both T&D losses and power plant fuel efficiencies. *It should also be noted that T&D losses and source energy savings calculations are often considered in the gross energy savings calculation instead of the net energy savings calculation*. In either case, savings should be reported with an indication of whether they include T&D losses and are based on source energy or end-use energy.

Other influences (in addition to free ridership, spillover, rebound, and T&D losses) that can determine net versus gross savings include:

- The state of the economy (recession, recovery, economic growth).
- Energy prices.
- Changes in facility operations (e.g., office building or hotel occupancy rates, changes in product lines or number of operating shifts in factors, or changes in thermostat settings or number of people living in homes). These are typically addressed in the gross savings analyses.

5.3 Approaches for Determining Net Savings

The following discussion presents the four approaches for determining the NTGR:

- **Self-reporting surveys.** Information is reported by participants and non-participants without independent verification or review.
- **Enhanced self-reporting surveys.** The self-reporting surveys are combined with interviews and documentation review and analysis.
- **Econometric methods.** Statistical models are used to compare participant and non-participant energy and demand patterns.
- **Stipulated net-to-gross ratios.** Ratios that are multiplied by the gross savings to obtain an estimate

of net savings and are based on historical studies of similar programs.

With respect to program size and scale, the two survey methods can be used with any program regardless of the number of participants. The third approach can only be used with programs with large numbers of participants because the models need large amounts of data to provide reliable results. The fourth approach can be used any time there is sufficient data to support a stipulated value.

In terms of timing, an NTGR analysis can be integrated into the gross impact analysis if the large-scale data analysis approach is used. With other gross impact analysis approaches, the NTGR is calculated independently, perhaps covering a longer period of time to more fully cover spillover and rebound effects. However, as with gross impact analysis, some of the approaches can be costly and evaluation resources can be limited. Accordingly, it is acceptable to perform NTGR analyses less frequently than the gross savings impact evaluation—perhaps every few years—as long as the market influences and participants' behavior are relatively consistent.

In terms of accuracy requirements, while econometric modeling can include tests for bias and precision and appropriate sample sizes can be determined, it is virtually impossible to define a precision target and a statistically valid sample size for the two self-reporting survey approaches. This challenge in surveying comes from the nature of collecting both qualitative and quantitative data from various participants and non-participants involved in the decision to install energy efficiency measures. In this case, evaluators attempt to survey all participants or intuitively select survey sample sizes.

The other uncertainty challenge in surveying is the subjective nature of assigning NTGRs to each participant. A participant is clearly a free rider if he or she would have installed the same project even if the program did not exist. Assigning NTGRs to individual participants is more complicated in cases where a participant *might* have installed the project, or would have installed it in two years if not for the program.

When non-participants are included in the NTGR analysis, care must be taken in selecting the appropriate comparison group. There is no single rule about what constitutes an appropriate comparison group, since the selection of the group depends on such factors as type of market transaction, survey methodology, and comparison purpose. The proposed non-participant comparison group and the criteria used in selecting this group should be discussed in the evaluation plan.

The following subsections briefly discuss the four approaches. (More information, specific to energy efficiency NTGR evaluations, can be found in CPUC, 2004.)

5.3.1 Self-Reporting Surveys

Survey-based stated intentions, or "self-reports," are a way to estimate free ridership by asking participants a series of questions on what they would have done in the absence of the program. Spillover estimates are developed and free ridership estimates are enhanced by non-participant surveys.

Surveys can be surprisingly complex to design and administer. They rely on respondent selection methods, survey instrument design, question wording, and implementation method to develop reliable results. One of the elements that should be addressed in surveys is self-selection bias. Self-selection bias is possible whenever the group being studied has any form of control over whether to participate: for example, people with strong opinions or substantial knowledge may be more willing to spend time answering a survey than those who do not. Self-selection bias is related to sample selection bias and can skew the results of an NTGR analysis that is not very well planned, funded, or executed.

Generally, the best use of self-reporting surveys has involved asking a series of questions with each question allowing a scale of responses. Surveys are either hard copy or Web-based instruments that are filled out by the interviewee, or perhaps conducted by phone with a professional surveyor (usually someone unfamiliar with energy efficiency). A typical initial question asked of participants is, "If the program had not existed, would you have installed the same equipment?" For a response, participants might choose between "definitely would have," "probably would have," "probably would not have," and "definitely would not have." This use of a scale, rather than a yes/no response, is thought to allow greater apparent confidence and precision in the estimate.

For free ridership, each of the responses is assigned a probability to determine the expected net savings. These estimates are then combined (additively or multiplicatively) into an individual participant free rider estimate. The participant estimates are subsequently averaged (or assigned a weighted average based on expected savings) to calculate the overall free ridership estimate. Similarly, non-participant responses are used to adjust a free ridership estimate and/or calculate spillover estimates.

Table 5-1 provides an example of a probability matrix used to determine a free ridership score. Note that the only 100 percent free ridership score is attained if a measure was already on order or installed prior to participation in the program. This approach was used in a process and impact evaluation of the Southern California Edison IDEEA program and an impact evaluation of the Energy Trust of Oregon's commercial and industrial programs.[1] (Note that the content of this table is intended only to illustrate the basic concepts.)

The survey approach is the most straightforward way to estimate free ridership and spillover. It is also the lowest-cost approach. It does, however, have its disadvantages in potential bias and with overall accuracy. For example, typical responses such as "don't know," missing data, and inconsistent answers are very hard to address without additional data collection. While there are ways to improve survey quality (e.g., using techniques like adding consistency check questions and adjusting the individual's estimate accordingly), the accuracy of simple self-reports is typically marginal.

5.3.2 Enhanced Self-Reporting Surveys

To improve the quality of NTGRs drawn from self-reported survey responses, the evaluation can rely on multiple data sources for the decision to install or adopt energy efficiency measures or practices. Some common additional data sources and techniques are:

- **Personal surveys.** Conducting in-person surveys is probably the best way to qualitatively improve the quality of self-surveys. Key participants in the decision to install efficiency measures can help determine the level of influence of the program on participants and non-participants. For commercial and government facilities, potential interviewees include managers, engineers, and facilities staff. Contractors, design engineers, and product manufacturers, distributors, and retailers can also provide information on the

Table 5-1. Example Free Rider Probability Assessment

Free-Ridership Score	Already Ordered or Installed	Would Have Installed Without Program	Same Efficiency	Would have Installed All of the Measures	Planning to Install Soon	Already in Budget
100%	Yes	Yes	—	—	—	—
0%	No	No	—	—	—	—
0%	No	Yes	No	—	—	—
50%	No	Yes	Yes	Yes	Yes	Yes
25%	No	Yes	Yes	Yes	No	Yes
25%	No	Yes	Yes	Yes	Yes	No
0%	No	Yes	Yes	Yes	No	No
25%	No	Yes	Yes	No	Yes	Yes
12.5%	No	Yes	Yes	No	No	Yes
12.5%	No	Yes	Yes	No	Yes	No
0%	No	Yes	Yes	No	No	No

Provided by Sami Khawaja of Quantec, LLC.

influences and motivations that determine the role of energy efficiency programs in the decision-making process. When working with professionals involved in the efficiency measure installation, individuals familiar with the program and projects should conduct the interviews. The interviewer should attempt to eliminate or at least minimize any bias they may have.

- **Project analysis.** This consists of two general types of reviews. The first is an analysis of the barriers to project installation and how the project addresses these barriers. The most common barrier is financial (project costs), so the common analysis is calculation of a project's simple payback. For example, if the project has a very short payback period without any program-provided benefits, then it may be considered as more likely to have been installed with or without the program.[2] The other type of analysis is reviewing any documentation the participant may have of the decision to proceed with the project. Such documentation may include internal memos or feasibility studies and can indicate the basis of the decision to proceed.
- **Non-specific market data collection.** Through the review of other information resources prepared for similar programs, the range of appropriate NTGRs can be estimated. Such resources might include analyses of market sales and shipping patterns, studies of decisions by participants and non-participants in similar programs, and market assessment, potential, or effects studies. Market sales methods rely on aggregate data on total sales of a particular technology in a given jurisdiction. They compare this sales volume with a baseline estimate of the volume that would have been sold in the absence of the program. The accuracy of these methods depends on the completeness and accuracy of the sales data, as well as the validity of the baseline estimate.

All or some of these three data sources can be combined with the written or Web-based participant and non-participant self-surveys to triangulate on an estimate of the free ridership, spillover, and rebound rates for that program.

Net-to-Gross Ratio Calculation Using Equipment Sales Data

In 1992 Baltimore Gas and Electric (BGE) offered a number of conservation programs, including a residential HVAC program. This program was designed to give consumers who were in the market to replace their HVAC systems incentives to choose a more energy-efficient heat pump or central air conditioner. BGE conducted an impact evaluation including a net-to-gross analysis designed to quantify the portion of energy-efficient HVAC purchases that could be attributed to BGE's program. Several sources of data were used:

- A survey of participants in BGE's residential HVAC program.
- Two surveys of customers who did not participate in BGE's residential HVAC programs.
- A survey of HVAC contractors who reported their sales of HVAC equipment by SEER (seasonal energy efficiency ratio).
- Data from the Air Conditioning and Refrigeration Institute that provided SEER levels for central air conditioners and heat pumps on an annual basis from 1981 through 1991.

These data provide a range of NTGRs from 0.74 to 0.92. An integrated approach provided what BGE considered the most reliable estimate:

Net-to-gross ratio = Net increase in purchases of qualifying equipment due to the program divided by the number of units sold under the program in 1992

$$= (28{,}300 - 18{,}700) \div 10{,}400$$
$$= 0.92$$

Thus, BGE concluded that an initial NTGR of 0.90 was appropriate.

Case study provided by Baltimore Gas and Electric.

5.3.3 Econometric Models

Econometric models, in this context, are mathematical tools that apply quantitative or statistical methods to the analysis of NTGRs. Econometric methods are sometimes considered the most accurate approach to calculating NTGRs when there are enough participants and truly comparable non-participants, and when the program is large enough to justify the cost of such analyses. The econometric models are closely related to, and can be the same models as, those described in Section 4.3 for calculating gross energy savings.

Various econometric methods have been used, with varying advantages and disadvantages. The models use energy (and demand) data from participants and non-participants over the same period to estimate the difference between gross savings (participant savings) and simple net savings (participant savings minus non-participant savings). The models differ in their mathematical and statistical calculation methods, but also in how they address complicating factors of bias that differentiate true NTGRs from simple comparisons of participant and non-participant savings. One particular element of surveying that econometric models attempt to address is self-selection bias.

5.3.4 Stipulated Net-to-Gross Ratio

This fourth approach, although not a calculation approach, is often used. NTGRs are stipulated in some jurisdictions when the expense of conducting NTGR analyses and the uncertainty of the potential results are considered significant barriers. In such a situation, a regulatory body sets the value, which is typically in the 80 to 95 percent range. Sources of stipulated NTGRs should be similar evaluations of other programs or, possibly, public utility commissions' requirements. Stipulated NTGRs should be updated periodically based on evaluations and review of other programs' calculated NTGRs.

5.4 Selecting a Net Savings Evaluation Approach

As mentioned in Chapter 4, selection of an evaluation approach is tied to the objectives of the program being evaluated, the scale of the program, the evaluation budget and resources, and specific aspects of the measures and participants in the program.

Another criterion—probably the most important—is cost. All four approaches can be used with any type of efficiency program, with the possible exception that the econometric modeling requires a program with a large number of participants. The lowest-budget approach is stipulated NTGR, followed by self-reporting surveys and enhanced surveys, and then econometric modeling (which incorporates the surveying activities). One option for keeping costs down while using the more sophisticated approaches is to conduct an NTGR analysis every few years and stipulate NTGRs for the intervening years.

5.5 Notes

1. Provided courtesy of Quantec, LLC.
2. Note that the need to decide when a consumer would have installed an energy project, based on the economic payback associated with a project, is an example of the subjective nature of free ridership. The choice of a specific payback period—2, 3, 4, etc., years—to define who is and who is not a free rider also has a subjective nature.

6: Calculating Avoided Air Emissions

Chapter 6 begins by describing two general approaches for determining avoided air emissions. It then presents several methods for calculating both direct onsite avoided emissions and reductions from grid-connected electric generating units. The chapter also discusses considerations for selecting a calculation approach.

6.1 General Approaches for Calculating Avoided Air Emissions

Avoided air emissions are determined by comparing the emissions occurring after an efficiency program is implemented to an estimate of what the emissions would have been in the absence of the program—that is, emissions under a baseline scenario. In practice avoided emissions are calculated with one of two different approaches: emission factor or scenario analysis.

1. **Emission factor approach**—multiplying the program's net energy savings (as determined by one or more of the approaches defined in Chapter 5) by an emissions factor (e.g., pounds of CO_2 per MWh) that represents the characteristics of displaced emission sources to compute hourly, monthly, or annual avoided emission values (e.g., tons of NO_X per year). The basic equation for this approach is (t = time period of analysis):

$$\text{avoided emissions}_t = (\text{net energy savings})_t \times (\text{emission factor})_t$$

2. **Scenario analysis approach**—calculating a base case of sources' (e.g., power plants connected to the grid) emissions without the efficiency program and comparing that with the emissions of the sources operating with the reduced energy consumption associated with the efficiency program. This is done with sophisticated computer simulation dispatch models and is usually only used with electricity saving programs. The basic equation for this approach is:

$$\text{avoided emissions} = (\text{base case emissions}) - (\text{reporting period emissions}$$

This chapter assumes that the net savings are calculated in a satisfactory manner, taking into account the issues raised in Section 3.8 with respect to quality of savings estimation, boundary areas, and additionality. Therefore, this chapter focuses on the various ways in which emission factors can be calculated and, for electricity efficiency programs, the basics of the scenario analysis approach. The first section of this chapter covers calculation of emission factors associated with avoided onsite fuel usage. The second section covers avoided emissions calculations for grid-connected electricity approaches—both emission factors and scenario analysis. The final section provides brief comments on selecting a calculation approach.

6.2 Direct Onsite Avoided Emissions

Direct, onsite avoided emissions can result when efficiency programs save electricity that would have been produced at a project site or when efficiency reduces the need for heat or mechanical energy, reducing onsite combustion of natural gas, fuel oil, or other fuels. Identifying the appropriate emission factor is fairly straightforward for onsite emissions such as those from residential or commercial combustion equipment, industrial processes, or onsite distributed generation. The emission factors are commonly calculated in one of two ways:

- **Default emission factors.** Default emission factors, available from standard resources, are based on the fuel and emission source being avoided. This is the most common approach and a wide variety of resources provide emission factors per unit of fuel consumption, including: manufacturer's equipment performance data, state-certified performance data, emission permit data, and generic emission data compiled by regulators or industry groups. Some data sources are the International Energy Agency (<http://www.iea.org>), Energy Information Agency (<http://www.eia.doe.gov>), and U.S. EPA (<http://www.epa.gov/ttn/chief/ap42/> and <http://cfpub.epa.gov/oarweb/index.cfm?action=fire.main>).

- **Source testing.** Source testing can determine the emission factors for a specific device (e.g., large-scale industrial boilers). Protocols for testing are available, but given the time and cost of such testing, this approach is usually only taken when required by environmental regulation. This may change if the value of avoided emissions makes source testing cost-effective as a part of a certification process, for example.

For direct onsite emissions, a typical emission factor is reported in units of emission per units of onsite fuel use. For example, a common CO_2 emission factor for natural gas is 117 pounds CO_2 per MMBtu. Such a value would be used with the quantity of avoided natural gas use to calculate emissions reductions, per the following equation:

avoided emissions = (net avoided natural gas use) × (emission factor)

For example, the following are calculations for a project that reduces natural gas consumption from a large industrial boiler by 10,000 MMBtu/year.

- Displaced steam use due to efficiency project = 10,000 MMBtu/year
  Steam boiler HHV efficiency = 80 percent
- Displaced natural gas usage = 10,000 MMBtu/yr ÷ 0.80 = 12,500 MMbtu/yr
- Avoided CO_2 emissions = 12,500 MMbtu/yr × 117 lbs CO_2/MMBtu = 1,462,500 lbs/yr
- Avoided emissions in tons = 1,462,000 lbs/yr ÷ 2,000 lbs/ton = 731 tons of CO_2/yr

The program evaluator must select onsite emission factors that provide sufficient accuracy to meet the goals of the evaluation. This requires selecting different emission factors for different time periods, places, and technologies. In addition, emission factors based on historical emission rates may need to be adjusted to account for new, more stringent regulations. Accounting for changing environmental regulation is an important consideration in calculating emission benefits.

Avoided Emissions from Combined Heat and Power Projects

Calculating the avoided emissions associated with a new combined heat and power (CHP) system involves special considerations. CHP systems generate both electricity and thermal energy from a common fuel source, and can be significantly more efficient than separate generation of electricity and thermal energy. In order to calculate the efficiency and the emissions impacts, one must compare the onsite energy use and emissions of the CHP facility to the combined onsite and grid energy use and emissions of the conventional systems. The onsite emissions can be calculated as described in Section 6.2. See Section 6.3 for how to calculate grid emissions.

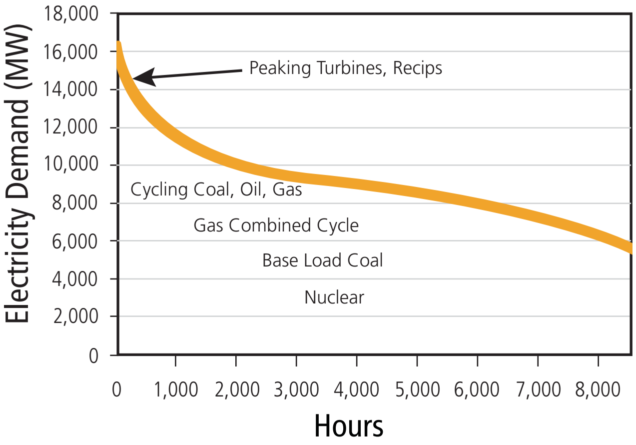
6.3 Emission Factors for Grid-Connected Electric Generating Units

Like the direct onsite case, emissions reductions from reduced electricity consumption occur because less fuel is combusted. However, calculating avoided grid emissions reductions is more complex because the fuel combustion in question would have occurred at many different existing or proposed electric generating units (EGUs), all connected to the electrical grid. Thus, emissions from displaced electricity depend on the dynamic interaction of the electrical grid, emission characteristics of grid-connected power plants, electrical loads, market factors, economics, and a variety of regional and environmental regulatory factors that change over time.

6.3.1 The Electricity Generation Mix

The electric grid is composed of a T&D system connecting a mix of generating plants with different emissions characteristics, which operate at different times to meet electricity demand. The mix of plants operating varies by region, and over time within regions—both as the demand changes from one hour to the next and as old plants are retired and new plants are built. A common way of looking at this varying generation mix is a load duration curve. The load duration curve shows the electricity demand in MW for a region for each of the 8,760 hours in the year. The hourly demand values are sorted from highest to lowest. Figure 6-1 shows an example from a typical eastern electric utility.

Figure 6-1. Load Duration Curve



Note: For illustrative value and not an actual system representation.

Provided by Joel Bluestone of EEA.

The figure shows that the highest hourly electric demand was 16,216 MW and the lowest demand was 5,257 MW. It also shows that the peaking turbines and reciprocating engines (recips) operated for only about 200 hours per year (in this case during very hot hours of the summer), while the base load coal and nuclear plants operated throughout the year. The total area under the curve is the generation needed to meet load plus line losses (in this case about 79.7 million MWh). The varying electric load is met with a large number of different types and sizes of generating units.

Figure 6-1 also indicates a typical mix of generating technologies. The generating units are dispatched based on a number of factors, the most important usually being the unit's variable cost—the cost of fuel, and operation and maintenance directly related to production. Base load units are run as much as possible because they are the least expensive; peaking and intermediate (cycling) units are used only when needed because of their higher costs. The type of units—base load, peaking, etc—that are the most "polluting" can vary from one region to another.

Compared to the base case, energy efficiency displaces a certain amount of generation during each hour that it operates. Efficiency essentially takes a "slice" off the top of the load curve for the hours that it occurs, displacing the last unit of generation in each of these hours. The displaced emissions can be estimated by multiplying the displaced generation by the specific emission rate of that unit or by preparing scenario analyses.

Depending on the hour of the day or year and the geographical location of the avoided electricity use, the displaced unit could be a cycling coal, oil, or steam unit; a combined cycle unit; a central station peaking turbine; or a reciprocating engine unit—or even a zero-emissions unit. The first challenge in calculating the avoided emissions for electricity generation is defining the mix of technologies displaced by the efficiency programs for the specific program location and during specific times of the year.

6.3.2 Calculating Avoided Emission Factors and Scenario Analyses

The methods for determining avoided emissions values for displaced generation range from fairly straightforward to highly complex. They include both spreadsheet-based calculations and dynamic modeling approaches with varying degrees of transparency, rigor, and cost. Evaluators can decide which method best meets their needs, given evaluation objectives and available resources and data quality requirements. Designers of programs or regulations that use these estimates may also wish to specify a method at the outset, and a process for periodic review of that method.

The emission rates of the electricity grid will vary over time. Thus, the emissions analyses are typically conducted annually for each year of the evaluation-reporting period for electricity saving programs. Emissions rates can also vary hour by hour as the mix of electricity plants operating changes to meet changing loads.

For natural gas and fuel oil programs, annual savings and hourly analyses are probably less critical. The decision to use an annual average analysis, an hourly analysis, or some time period of analysis in between is up to the evaluator to decide based on evaluation objectives and available resources.

6.3.3 Emission Factors Approach

This section describes two methods for calculating avoided emission factors:

- Calculating emissions rates using a simple "system average" displaced emissions rate obtained from an emissions database. This generally produces less precise estimates.
- Calculating emissions rates using a "medium effort" calculation method, such as estimating regional or state average emission rates for marginal generators or matching capacity curves to load curves. This generally results in moderately precise avoided emission estimates.

Section 6.3.3 further describes these two methods, beginning with a discussion about approaches for calculating emission factors for new and existing power plants.

Operating and Build Margin Emissions Rate

The load duration curve in Section 6.3.1 depicts an existing generation mix. However, efficiency could also prevent the need for future power plant construction. Even if gas-fired generation is currently what is avoided, energy efficiency can avoid the construction of a new power plant, such that emissions from that plant will be avoided as well. For most energy efficiency program activity in the United States, it is safe to assume that only existing generator emissions are avoided in the short term of two to five years. However, if the analysis is estimating impacts over a longer period of time and/or the scale of the programs being evaluated is large enough, then new units could be considered as well.

The emission factor from a generating unit that would not be run due to energy efficiency is called the operating margin (OM). The emission factor from a generating unit that would not be built is called the build margin (BM). In general terms, avoided emissions can be estimated by determining the extent to which an efficiency program or portfolio affects the BM and OM and either (a) determining appropriate emission factors for the BM and OM using the emission factor approach or (b) accounting for new and existing generating units when using the base case and efficiency scenario approach.

The general formula for calculating emission rates for determining avoided emissions rates is:

$$ER = (w) \times (BM) + (1 - w) \times (OM)$$

where: ER is the average emission rate (e.g., tons of CO_2-equivalent / MWh)

BM is the build margin emission factor (e.g., t CO_2-equivalent / MWh)

OM is the operating margin emission factor (e.g., t CO_2-equivalent / MWh)

w is the weight (between 0 and 1) assigned to the build margin

Time is explicit in this equation. That is, the emissions reduction can vary from year to year (or in theory from hour to hour) as the variables w, BM, and OM change over time. In this formula, w indicates where the generation produced (or reduced) by the project activity would have come from in the baseline scenario. A weight of 1 means that all generation produced or saved by the project activity would have come from an alternative type of new capacity built in place of the project activity (the BM). A weight between 0 and 1 means that some of the generation would have come from new capacity (BM) and the remainder from existing capacity (the OM). A weight of 0 means that all of the generation would have been provided by existing power plants, and no new capacity would have been built in place of the project activity.

One approach to determining OM and BM can be found in the WRI/WBCSD *Protocol Guidelines for Quantifying GHG Reductions from Grid-Connected Electricity Projects* (see <http://www.wri.org/climate>). In this approach, there are three options for selecting the BM emission factor:

- **Option #1. Use a project-specific analysis to identify the type of capacity displaced.** Under this option, the BM emission factor is representative of a single type of power plant. This type of power plant will be either (a) the baseline candidate (i.e., baseline power plant) with the lowest barriers or greatest net benefits or (b) the most conservative, lowest-emitting baseline candidate.
- **Option #2. Use a conservative "proxy plant" to estimate BM emissions.** Under this option, the BM emission factor is determined by the least-emitting type of capacity that might reasonably be built. In some cases, this baseline candidate could have an emission rate of zero (e.g., renewables). Another way to determine a proxy is to look at the plants that have recently been built and connected to the grid.
- **Option #3. Develop a performance standard to estimate the BM emission factor.** Under this option, the BM emission factor will reflect a blended emission rate of viable new capacity options.

If the BM is included in the analyses, it must be explicitly specified, including the basis for its calculation. In recent years, estimates for BM emission rates have been based on advanced-technology coal plants or gas-fired, combined-cycle power plants, as most new plants adopt this technology. However, with new technologies being developed and renewable portfolio standards becoming more prevalent, changes in market conditions should be tracked and accounted for when using a BM emission factor.

System Average Emissions Rate

One simple approach for calculating emissions reductions from efficiency programs is to use regional or system average emission rates. Determining a system average rate involves dividing total annual emissions (typically in pounds) from all units in a region or power system (i.e., within the relevant grid boundary) by the total energy output of those units, typically in MWh.

Sources for average emissions rates include the Ozone Transport Commission's "OTC Workbook" (OTC, 2002), the Clean Air and Climate Protection Software (ICLEI, 2003), and EPA's eGRID database (EPA, 2007). Each of these tools contains pre-calculated emissions rates averaged at the utility, state, and regional level. The rates vary by time period, dispatch-order, and region, as discussed further in the medium-effort discussion below. A shortcoming of this approach is that it does not account for the complexity of regional power systems. While the tools above offer a wide variety of emission factors, it can be difficult to select the most appropriate approach. In many regions, the marginal units displaced by energy efficiency programs can have very different emissions characteristics from the base load units that dominate the average emissions rate.

Another shortcoming of this approach is that energy efficiency savings tend to vary over time, such as savings from an office lighting retrofit that only occurs during the workday. Using an annual average emission factor that lumps daytime, nighttime, weekday, and weekend values together can skew the actual emissions benefits calculation.

A system average emission rate may be purely historical, and thus fail to account for changing emissions regulations and new plant additions. Historical system averages will tend to overestimate emissions impacts if emissions limits become more stringent over time. Alternatively, a system average emissions rate could be estimated for a hypothetical future system, based on assumptions about emissions from new plants and future regulatory effects on existing plants.

In summary, this is an easy approach to apply but the tradeoff can be relatively high uncertainty.

Medium Effort Calculation Approaches

Between system average calculations and dispatch modeling (scenario analysis) lie several "medium effort" approaches to estimating displaced emission rates. These methods have been developed to provide a reasonably accurate estimate of displaced emissions at a lower cost than dispatch modeling. They typically use spreadsheets and require compilation of publicly available data to approximate the marginal generating units supplying power at the time that efficiency resources are reducing consumption. The two major steps in a spreadsheet-based analysis are determining the relevant set of generating units (accounting for the location of the efficiency program's projects, as well as transfers between the geographic region of interest and other power areas) and estimating the displaced emissions from those units. The following approaches indicate how "medium effort" emission factors can be determined.

- **Estimating grid average emission rates for marginal generators.** This approach assumes that total emissions are reduced at an average emission rate for each additional kWh of energy reduction (a significant simplification for efficiency activities). To more precisely estimate the impact on the marginal generators that are most likely to be displaced, regional or state average rates are adopted that exclude the baseload generators not "backed off" by efficiency programs.[1] The downside of this approach is that it does not capture the precise subset of generators actually following load and thus subject to displacement. The emission rates of load-following units can vary significantly from the overall regional average for marginal generators. While the eGRID database is based on historical data, expected new units can also be added to this type of calculation. This approach was adopted in a 2006 analysis of New Jersey's Clean Energy Program, (see U.S. DOE, 2006).
- **Matching capacity curves to load curves.** As discussed above, generating units are typically dispatched in a predictable order based on cost and other operational characteristics. This means it is possible, in principle, to predict which unit types will be "on the margin" at a given load level, and thereby predict the marginal emission rates. Data on regional power plants may be used to develop supply curves representing different seasons and times of day. These curves are then used to match regional electricity loads to characteristic emission rates. Although this method may be able to use readily available public data, it is based on a simplified view of dispatch process that does not account for transmission congestion.

Existing vs. New Generating Units

The three approaches for calculating an emission factor are all ways to estimate avoided grid electricity emissions, given a defined current or future set of electricity generating units. If energy efficiency is assumed to reduce the need for new generation, a complementary type of computer modeling may be useful: power sector forecasting and planning models. Also, some integrated national energy models such as NEMS and IPM estimate both, calculate future changes in generating units and also providing an assessment of how generation would meet load. Such models can represent the competition between different types of generators, adding new generating capacity to meet load growth within the constraints of current and anticipated future environmental regulations and emission trading programs. This type of model addresses both the environmental regulatory effects and addition of new generating units in an integrated fashion.

Like system average methods, these methods do not provide an approach to determine how large a geographic region should be considered or how interregional transfer is estimated. However, both of them improve upon the system average with respect to identifying which generators are marginal. In either case, the analysis must include the effect of changing environmental regulation, as discussed above.

A significant advantage of using time-varying emission rates, either from dispatch models or other approaches, is that they can match up to the time-varying savings from efficiency programs. Even if an hour-by-hour load shape is not used, at least having seasonal weekday and weekend and nighttime and daytime values (i.e., six emission factors) to match up the equivalent time period net efficiency savings will significantly improve estimates over the other emission factor methods described above.

6.3.4 Scenario Approach

At the other end of the complexity spectrum from calculating simple emission factors, computer-based "hourly dispatch" or "production cost" models capture a high level of detail on the specific EGUs displaced by energy efficiency projects or programs.[2] The models are used to generate scenarios of the electric grid's operation, with and without the efficiency program being evaluated. A scenario analysis can estimate avoided emissions much more precisely than the emission factors methods described above. As such, it is a preferred approach where feasible.

An hourly dispatch model simulates hourly power dispatch to explicitly estimate emissions from each unit in a system. That system can represent the current grid and generating units, or can represent an anticipated future system based on detailed assumptions about additions, retirements, and major grid changes. However, dispatch models do not model the competition among different generating technologies to provide new generation. In general, the model produces a deterministic, least-cost system dispatch based on a highly detailed representation of generating units—including some representation of transmission constraints, forced outages, and energy transfers among different regions—in the geographic area of interest.

If the power system is altered through load reduction or the introduction of an efficiency program, the model calculates how this would affect dispatch and then calculates the resulting emissions and prices. The basis for the scenario approach is that a dispatch model is run with and without the efficiency program and the resulting difference in emissions is calculated. The models can also be used to provide hourly, monthly, or annual emission factors.

With a dispatch model, base case data are either (a) inputted from historical dispatch data provided by utilities or a system operator or (b) modeled on a chronological (hourly) basis.[3] The model is then run with the new efficiency resource to obtain the "efficiency case." Commercial models are typically sold with publicly available data already entered, including planned capacity expansions. Dispatch modeling is the most precise means of quantifying avoided emissions, because it can model effects of load reductions that are substantial enough to change dispatch (as well as future changes such as new generating units or new transmission corridors) on an hourly basis, taking into account changes throughout the interconnected grid.

On the downside, dispatch can be labor-intensive and difficult for non-experts to evaluate. These models can also be expensive, although the costs have been reduced over recent years and—particularly if the results can be applied to a large program or several programs—the improved estimate can be well worth the incremental cost. Accordingly, they are probably most appropriate for large programs or groups of programs that seek to achieve significant quantities of electrical energy efficiency or long-term effects. For large statewide programs, the modeling costs may be relatively small compared to the program and evaluation costs; CPUC, for example, is currently using dispatch modeling to determine the avoided greenhouse gases from various efficiency portfolios.[4]

6.4 Selecting an Approach for Calculating Avoided Emissions

The choice of evaluation approach is tied to the objectives of the program being evaluated, the scale of the program, the evaluation budget and resources, and the specific emissions the program is avoiding. For direct onsite fuel savings and the resulting avoided emissions, standard emission factors can be used. This is a common practice, except perhaps for very large industrial individual efficiency projects.

For electricity savings programs, system average emission values can be used but should be avoided except in the simplest estimates. There are also medium effort methodologies that can fairly accurately quantify the effects of electricity energy efficiency programs. However, the most precise approaches involve dispatch modeling and the resulting detailed calculation of hourly emissions. While the costs and complexity of these models has limited their use in the past, this is beginning to change. Dispatch models are potentially cost-effective evaluation tools that should be considered for evaluations of large-scale programs.

6.5 Notes

1. The latest version of EPA's eGrid database (U.S. EPA, 2007) includes one such calculation.
2. These models are also called "production cost models."
3. Historical data can be used to calibrate the chronological model. However, using historical data directly for the base year can lead to results that include unusual system performance during the base year as well as changes from the efficiency program(s).
4. See <http://www.ethree.com/cpuc_ghg_model.html>.

Wisconsin Focus on Energy Program's Calculation of Avoided Emissions

Evaluators for Wisconsin's Focus on Energy public benefits energy efficiency program have estimated emission factors for the plants serving Wisconsin and used these data to estimate long-term avoided emissions associated with the Focus programs. The evaluation team developed a model to estimate the generation emission rates for NO_X, SO_X, CO_2, and mercury using hourly measured emissions data from EPA for the power plants supplying Wisconsin (using a medium effort calculation approach). Emission factors from reduced use of natural gas at the customer site were also taken from EPA data.

Using the marginal cost emission rates and evaluation-verified gross electricity savings estimates, the Focus programs together potentially avoided 2,494,323 pounds of NO_X; 4,107,200 pounds of SO_X; over 2,369 million pounds of CO_2; and over 15.9 pounds of mercury from inception to December 31, 2006 (See Table 2-11 of the Focus on Energy Public Benefits Evaluation *Semiannual Report—FY07, Mid-Year*, May 10, 2007).

To complement this effort, Wisconsin's Department of Natural Resources (DNR) has developed an emissions registry to track emissions reductions in Wisconsin. The ongoing reporting of emissions reductions associated with Focus programs' energy impacts has been the basis for entries to DNR's Voluntary Emissions Reduction Registry (<http://www.dnr.state.wi.us/org/aw/air/registry/index.html>).

For use with the registry, the Focus on Energy evaluator provides independent third-party verification for one of the residential programs, ENERGY STAR Products. This program promotes the installation of energy-efficient appliances, lighting, and windows. Drawing upon the evaluation activities conducted over the past four years, the emissions savings from the Energy Saver compact fluorescent light bulb portion of the program were verified for the Registry. The calculations, assumptions, and research activity backup that supports the registered reductions in emissions associated with the evaluated energy impacts of the program are cited and available on the state's DNR Web site.

It should be noted that Wisconsin's power plants are included in the federal SO_2 cap and trade program (acid rain provisions). In this cap and trade system, SO_2 emissions may not be considered reduced or avoided unless EPA lowers the SO_2 cap. One can say that the program avoided generation that previously emitted SO_2, but one cannot claim that future SO_2 emissions will actually be reduced due to the effect of the trading program. Starting in 2009, the plants will also be subject to a cap and trade program for NO_X (the Clean Air Interstate Rule), which will have the same effect.

Provided by David Sumi of PA Consulting Group.

7: Planning an Impact Evaluation

Chapter 7 builds on preceding chapters and presents the steps involved in planning an impact evaluation. These include the development of evaluation approaches, budgets, and a schedule. The first section discusses how evaluation planning and reporting is integrated into the program implementation process, while the second section presents seven key issues and questions that help determine the scope and scale of an impact evaluation. The last section provides guidance on preparing an evaluation plan and includes model outlines and checklists for conducting an evaluation plan.

7.1 Integration of Evaluation into the Program Implementation Cycle

After reading this chapter, and this Guide, the reader should have the background needed for preparing an evaluation plan to document gross and net energy and demand savings, and avoided air emissions from an energy efficiency program. However, this Guide cannot be a substitute for the experience and expertise of professional efficiency evaluators. While it can be used in preparing an evaluation plan, it may be best used to oversee the evaluation process as implemented by professional evaluators, whether they be internal staff or outside consultants.

Before describing the evaluation planning process, it is important to understand how it is integral to what is typically a cyclic planning-implementation-evaluation process. In most cases the overall cycle timeframe is consistent with program funding and contracting schedules.

These program implemetation cycles can be one or two years, or even longer. The point at which programs are being designed is when the evaluation planning process should begin. This is primarily so that the program budget, schedule, and resources can properly take into account evaluation requirements.[1] It is also a way to ensure that data collection required to support expected evaluation efforts is accommodated at the time of implementation.

The Program Implementation Cycle

Evaluation results are used to make informed decisions on program improvements and future program designs and offerings. The program implementation cycle is one in which programs are designed, then implemented, and then evaluated. Using the results of the evaluation, programs are re-examined for design changes and then modified so that those design changes result in improved program implementation efforts. This cycle provides for a continuing process of program improvement, so that the programs match available market opportunities and continually improve their cost-effectiveness over time.

Source: CPUC, 2004.

Evaluations should be completed within a program cycle, so that evaluation results can not only document the operations and effects of the program in a timely manner, but also provide feedback for ongoing program improvement, provide information to support energy efficiency portfolio assessments, and help support the planning for future program cycles. For impact evaluations that examine energy savings of certain measures and program mechanisms, the evaluation information can also be used to inform future savings estimates and reduce future evaluation requirements and costs.

Figure 7-1. Program Implementation Cycle with High-Level Evaluation Activities

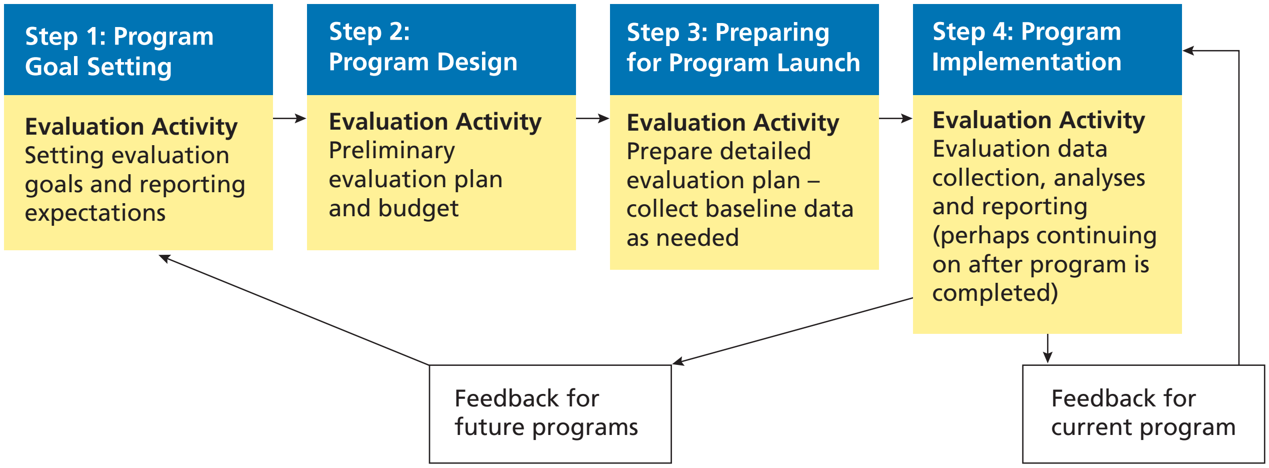


Figure 7-1 shows the energy efficiency program implementation cycle, emphasizing evaluation activities, as well as feedback to the current and future programs.

The steps displayed in Figure 7-1 are further described below:

- **Program goal setting.** When a program is first envisioned, often as part of a portfolio of programs, is when both program goals and evaluation goals should be considered. If the program (or portfolio) goal is to save electricity during peak usage periods, for example, the evaluation goal can be to accurately document how much electricity is saved during the peak (gross impact) and how much of these savings can be attributed to the program (net impact).
- **Program design.** Program design is also when the evaluation design effort should begin. The objective should be a *preliminary evaluation plan and budget*. The seven issues described in Section 7.2 should be raised, although not necessarily fully addressed, at this time. Whereas a program design is usually completed at this stage, it is likely that the evaluation plan will not be fully defined. This is typically because of the iterative nature of integrating the program design and evaluation process and the timing for when the evaluator is brought into the team. It is not unusual, although not always best practice, to select the evaluator after the program has been designed. Regardless, specific evaluation goals and objectives should be set and priorities established based on factors including perceived risks to achieving the savings objectives.
- **Preparing for program launch.** Program launch is when activities, program materials, and timing strategies are finalized and made ready, contracts (if needed) are negotiated, trade allies and key stakeholders are notified, and materials and internal processes are developed to prepare for program introduction and implementation. The detailed evaluation plan should be prepared before the program is launched—or if not, soon after it is launched. (An outline of such a plan is presented in Section 7.3.) It is in this plan that the seven evaluation issues are fully addressed and resolved, including specifying the data needed to perform the evaluation.

This is also the time when some baseline data collection can take place. A major reason for starting the evaluation planning process well before a program is launched is if baseline data collection is required.

The overall evaluation plan should be reviewed with program implementers and possibly with the appropriate oversight body or bodies to ensure that it meets the information needs of policy-makers, portfolio managers, and/or regulators. This is also the time when evaluation staff or consultants are assigned to the program evalu-

ISO-NE M&V Manual for Wholesale Forward Capacity Market (FCM)

In 2007, the Independent System Operator of New England (ISO-NE) developed an M&V Manual that describes the minimum requirements the sponsor of a demand resource (DR) project must satisfy to qualify as a capacity resource in New England's wholesale electricity forward capacity market (FCM). DRs eligible to participate in FCM include demand response, emergency generation, distributed generation, load management, and energy efficiency. DRs are eligible to receive a capacity payment ($/kW per month) based on the measured and verified electrical reductions during ISO-specified performance hours. The Manual was developed with input from key stakeholders in the region, including members of the New England Power Pool, ISO-NE, the New England state regulatory staff, electric utility program administrators, Northeast Energy Efficiency Partnerships, and energy service, consulting and technology providers. The Manual specifies the minimum requirements a project sponsor's M&V Plan must address, including:

- **M&V methods.** The sponsor must choose from options based on the IPMVP Options A through D (or equivalent). Other M&V techniques may be used in combination with one or more of these, including engineering estimates supplemented with data collected on the equipment affected by the measures, and/or verifiable measure hourly load shapes (which must be based on actual metering data, load research, or simulation modeling). All DR including distributed generation and emergency generation must be metered at the generator.
- **Confidence and precision.** The project sponsor must describe a method for controlling bias (e.g., calibration of measurement tools, measurement error, engineering model) and achieving a precision of +/- 10 percent, with an 80 percent confidence level around the total demand reduction value. This requirement also applies to precision level for statistical sampling.
- **Baseline conditions.** The Manual specifies baseline condition requirements for failed equipment (codes/standards or standard practice, whichever is more stringent), early retirement (codes/standards or measured baseline), and new construction (codes/standards or standard practice). Where standard practice is used, baseline conditions must be documented and meet the confidence and precision requirements. For distributed generation and emergency generation, the baseline is zero. The baseline for real time demand response is calculated using a modified rolling average of the host facility load on non-event weekdays during the same hours as the called event.
- **Measurement equipment specifications.** The project sponsor must describe measurement, monitoring, and data recording device type that will be used (and how it will be installed) for each parameter and variable. Any measurement or monitoring equipment that directly measures electrical demand (kW) (or proxy variables such as voltage, current, temp. flow rates, and operating hours) must be a true RMS measurement device with an accuracy of at least ±2 percent.
- **Monitoring parameters and variables.** The project sponsor must describe variables that will be measured, monitored, counted, recorded, collected, and maintained, and meet minimum requirements for data to be collected by end-use and monitoring frequency.

Provided by Julie E. Michals, Public Policy Manager, Northeast Energy Efficiency Partnerships, Inc. For more information, see Final M&V Manual (dated 4/13/07) posted on ISO's Web site at: <http://www.iso-ne.com/rules_proceds/isone_mnls/index.html>.

---

ation. Issues for selecting evaluators are discussed in Section 7.2.7.

- **Program implementation.** This is when the evaluation actually occurs. Some baseline and all the reporting period data are collected, the analysis is done, and the reporting is completed. Given the often retrospective nature of evaluation, the evaluation activities can actually carry on after the program implementation step is completed.

Closing the Loop—Integration of Implementer and Evaluator

There has been a noticeable paradigm shift in evaluation in recent years. The old model brought in the evaluator on the tail end of the project to assess delivery, cost-effectiveness, and achievement of stated goals. In most cases, the evaluator was faced with the challenge of having to conduct analysis with less than perfect data. Even when data were available, the evaluator may have revealed facts that would have been useful early in the process for making course corrections. Had these corrections been made, better services would have been delivered. A different model brings the evaluator in at the onset of the program, as an integral part of the team. Program goals are linked to specific metrics, which are linked to specific data collection methods. The evaluator can provide feedback in real time to provide instant assessment and determination that the correct course is being followed. This model needs to be balanced with the possible conflicting nature of evaluation goals—the implementer's goal of understanding and improving the program performance and a regulating authority's goal of ensuring that the savings reported are "real." This conflict is more likely to occur if the implementer and evaluator are independent entities, as commonly required to meet the regulator's goal.

Although it is preferable to start an evaluation prior to the program launch in order to collect baseline information, in most cases the evaluation and program start simultaneously due to the common interest in initiating a program as soon as possible. Thus, activities to support data collection usually begin after the program is up and running, and hopefully early enough in the program cycle to provide feedback and corrective recommendations to program implementers in time for the program to benefit from those recommendations. In addition, impact evaluation activities can support program progress tracking, such as measure installation tracking and verification.

In terms of reporting, evaluation information can be summarized and provided on any time cycle. The key is to get the information needed to implementers so they can adjust existing programs and design new ones using current and relevant information. The evaluation activities may be conducted with oversight bodies providing review and approval so that specific reporting requirements may be necessary.

For future program designs, ex ante savings estimates can be adjusted based on program evaluation results. Assumptions underlying the efficiency potential analysis used at the beginning of the program cycle for planning can then be updated based on the full net impact analysis. These data then feed back into the goal setting and potentials analysis activities, and the cycle repeats to allow for an integrated planning process for future programs.

7.2 Issues and Decisions That Determine the Scope of an Impact Evaluation

Numerous elements and decisions go into the design of an impact evaluation, but there are seven major issues that require some level of resolution before the budget and the evaluation plan are prepared:

1. Define evaluation goals and scale (relative magnitude or comprehensiveness).
2. Set a time frame for evaluation and reporting expectations.
3. Set a spatial boundary for evaluation.
4. Define a program baseline, baseline adjustments, and data collection requirements.
5. Establish a budget in context of information quality goals.
6. Select impact evaluation approaches for gross and net savings calculations and avoided emissions calculations.
7. Select who (or which type of organization) will conduct the evaluation.

These issues are presented in what can be considered a linear sequence, but many are interrelated and the overall planning process is certainly iterative. The end result of addressing the above seven issues is an evaluation plan. Experience has indicated that, if the funding and time requirements for reliable evaluations are not fully understood and balanced with information needs and accuracy expectations, efforts can be under-supported and fail to provide the results desired.

7.2.1 Defining Evaluation, Goals, and Scale

This subsection helps the evaluation planner define evaluation goals, the overall scale of the effort, the specific program benefits to be evaluated, and whether any other evaluations will be concurrently conducted and coordinated.

Evaluations should focus on a program's performance at meeting its key goals and, if desired, provide information for future program planning. To this end, program managers and regulators need to be assured that the evaluations conducted will deliver the type and quality of information needed. Under-designed evaluations can waste valuable resources by not reliably providing the information needed or delay the start of an evaluation. Delays can make it impossible to collect valuable baseline data and postpone the results so that they cannot be used for current program improvement or future program design.

Evaluations can also be over-designed, addressing issues that are not priority issues or employing methods that could be replaced by less costly approaches. There is a need to prioritize evaluation activities so that evaluation resources—typically limited—can be focused on the issues of importance. Like many activities, an evaluation that is well-defined and affordable is more likely to be completed successfully than one with undefined or unrealistic objectives and budget requirements.

Setting goals involves defining evaluation objectives and the specific information that will be reported-out from the impact evaluation. The scale of the evaluation is more of a subjective concept, indicating how much effort (e.g., time, funding, human resources) will be expended on the evaluation.

Program Objectives and Information Reporting

As discussed in the beginning of this Guide, evaluations have two key objectives:

1. Document and measure the effects of a program in order to determine how well it has met its efficiency goals with respect to being a reliable, clean, and cost-effective energy resource.
2. Understand why those effects occurred and identify ways to improve current and future programs.

Additional objectives of evaluation can include determining the cost-effectiveness of a program and, when public or ratepayer funds are involved, documenting compliance with regulatory requirements. One of the other potential objectives of the impact evaluation effort is to provide policy makers and portfolio decision-makers with the information they need to identify programs to run in the future and assess the potential savings from these programs.

Therefore, the first step in planning an evaluation is simply picking which of these objectives are applicable and making them more specific to the evaluated program. Some typical impact evaluation objectives are:

- Measure and document energy and peak savings.
- Measure and document avoided emissions.
- Provide data needed to assess cost-effectiveness.
- Provide ongoing feedback and guidance to the program administrator.
- Inform decisions regarding program administrator compensation and final payments (for regulated programs and performance-based programs).
- Help assess if there is a continuing need for the program.

In practice, the selection of objectives will be shaped by many situational factors. Among the most important are:

- Program goals—it is also important that *program* goals must be quantifiable and able to be measured.

How the goals will be measured (evaluated) must also be taken into account in the program planning process.

- Whether the program is a new effort, an expanding effort, or a contracting effort.
- The policy and/or regulatory framework in which the evaluation results will be reported.
- The relative priority placed upon the evaluation's comprehensiveness and accuracy by the responsible authorities (i.e., the budget and resources available).

In terms of reporting out impact evaluation results, the key parameters are the units and time frame. Some examples are:

- Electricity savings: kWh saved per year and per month.
- Demand savings (example 1): kW saved per month of each year of program, averaged over peak weekday hours.
- Demand savings (example 2): kW savings coincident with annual utility peak demand, reported for each year of the program.
- Avoided emissions (example 1): metric tons of CO_2 and SO_X avoided during each year of the program.
- Avoided emissions (example 2): metric tons of NO_X avoided during ozone season months of each year of the program.
- Lifetime savings (savings that occur during the effective useful life of the efficiency measure): MWh saved during measure lifetime, in years.

In addition, as discussed in Section 3.4 and Appendix D, evaluation results, like any estimate, should be reported as expected values with an associated level of variability.

Evaluation Scale

Scale refers to an evaluation effort's relative magnitude or comprehensiveness. Will it be a major effort, a minor effort, or something in between? The following are some attributes that set the scale of an evaluation. The scale can be translated into resource requirement (time, cost, equipment, and people) estimates. Understanding the requirements and comparing them with the objectives, and resources available, should result in a well-balanced evaluation effort.

- How large is the program in terms of budget and goals? Larger programs tend to have larger evaluations.
- Is it a new program with uncertain savings or an established program with well-understood savings? Established programs with a history of well-documented savings may not require the same level of evaluation that a new program, with no history, requires. Related to this consideration is how much confidence exists in pre-program (ex ante) savings estimates. If a fair amount of effort has gone into feasibility studies and perhaps pre-testing, then less of an evaluation effort may be required.
- Is the program likely to be expanded or contracted? A program that may be expanded (i.e. increased in budget) probably deserves more analyses to confirm if it should be expanded than one that is not likely to receive additional funding or may even be cancelled.
- How accurate and precise an estimate of energy and demand savings is required? Less uncertainty generally requires bigger budgets. On one end of the uncertainty scale is simply verifying that the individual projects in a program were installed (and using deemed savings to determine savings). On the other end are rigorous field inspections, data collection, and analyses on all or a large sample of projects in a program.
- Do savings need to be attributed to specific projects within a program? If savings values for each project are desired, then a census evaluation is required. This is more costly than evaluating a sample of projects.
- How long, typically in years, does the evaluation need to be conducted? Typically, longer evaluation cycles require more funding.

- What is the time interval for reporting savings? For example, reporting annual or monthly savings estimates is usually much simpler than reporting hourly savings. This is particularly important when deciding how accurate an estimate of demand savings needs to be. As discussed in Chapter 3, there are different ways to calculate and report demand savings, with very different levels of effort required.

- What are the reporting requirements and who must review (and approve) evaluation results? While all evaluations should have well-documented results, the frequency that savings need to be reported, and to what audience—for example, a regulatory body—can influence the scale of the effort.

- Are avoided emissions also to be determined, and will the avoided emissions benefits be used in a regulatory program? As discussed in Chapter 6, emissions can be calculated simply or with significant effort and accuracy. If avoided emissions values will be used in a regulated program, the analyses may be subject to specific requirements and third-party verification.

- Are other co-benefits to be evaluated and quantified? If this is more than an anecdotal exercise, then additional resources will be required.

Other Evaluation Efforts and Other Programs

While this Guide is focused on impact evaluations, there are other types of evaluations (as described in Chapter 2 and Appendix B). If other evaluations, such as process or market effects evaluations, are to be conducted, their plans should be integrated with the impact evaluation plan. If cost-effectiveness analyses are to be conducted, it is critical to define which cost-effectiveness test(s) will be used and thus what impact evaluation data are needed. Furthermore, if more than one program is being evaluated and the programs may have some interaction, then coordination of the programs, their evaluations, and the assigning of net benefits to one program versus another need to be coordinated.

Evaluating Co-Benefits

This Guide is focused on documenting three categories of impacts or benefits associated with energy efficiency programs: energy savings, demand savings, and avoided air emissions. However, as discussed in Chapter 3, there are other potential benefits of energy efficiency. As part of the planning process, it must be decided which of these benefits, if any, will be evaluated and how.

7.2.2 Setting the Time Frame for Evaluation and Reporting

This subsection helps the evaluation planner define when the evaluation effort will start, how long it will last, for what time segments and intervals the savings data will be collected and reported (granularity), and the point at which evaluation reports will be available.

The evaluation time frame has two components:

1. **When and over what period of time the evaluation effort will take place.** A standard evaluation would begin before the start of the program implementation (to collect any baseline data) and continue for some time after the program is completed to analyze persistence of savings. However, the actual timing of the evaluation is influenced by several, often competing, considerations. These considerations include:
   a. What will be the time period of analyses, i.e. how many years?
   b. Will persistence of savings be determined, and if so, how?
   c. The timing for policy decisions and evaluation planning.
   d. The desire to have early feedback for program implementers.
   e. Program lifecycle stage (evaluating a first time program or a long-established program)
   f. Evaluation data collection time lags.
   g. Regulatory and/or management oversight requirements.
   h. Contract requirements for reporting savings for “pay for performance” programs.

i. Timing requirements to use the evaluation results to update specific measure energy and demand savings, and measure life estimates.

j. Reporting requirements—whether a single final program report is needed or whether interim or evenly monthly reports are desired.

2. **The time granularity of evaluation analyses.** This relates to whether 15-minute, hourly, monthly, seasonal, or annual data collection and savings reporting are required. The granularity decision is based on the uses of the information from the evaluation. Annual savings data are generally only useful for an overview of the program benefits. More detailed data are usually required for both cost-effectiveness analyses and resource planning purposes. For avoided emissions, annual values are typical; however, for certain programs, such as smog programs, there are specific seasons or time periods of interest.

   If demand savings are to be calculated, the choice of definition (e.g., annual average, peak summer, coincident peak, etc.) is related to time granularity. Chapter 3 includes a discussion of the different definitions and describes how this decision greatly influences the data collection requirements and thus the effort required to complete the evaluation.

7.2.3 Setting the Spatial Boundary for Evaluation

This subsection helps the evaluation planner define the assessment boundary, in at least general terms, for the evaluation.

When evaluating energy, demand, and emission savings, it is important to properly define the project boundaries (i.e., what equipment, systems, or facilities will be included in the analyses). Ideally, all primary effects (the intended savings) and secondary effects (unintended positive or negative effects), and all direct (at the project site) and indirect (at other sites) avoided emissions will be taken into account.

From a practical point of view, and with respect to energy and demand savings, the decision concerns whether savings will be evaluated for specific pieces of equipment (the "boundary" may include, for example, motor savings or light bulb savings), the end-use system (e.g., the HVAC system or the lighting system), whole facilities, or even an entire energy supply and distribution system. For avoided emissions calculations, the boundary assessment issues are discussed in Section 3.8.

7.2.4 Defining Program Baseline, Baseline Adjustments, and Data Collection Requirements

This subsection helps the evaluation planner define whether a project- or performance-based baseline will be used and decide on the basis for quantifying the baseline (e.g., existing equipment performance, industry typical practice, minimum equipment standards), which major independent variables will be considered in the analyses, and what data will need to be collected to analyze benefits.

As mentioned before, a major impact evaluation decision is defining the baseline. The baseline reflects the conditions, including energy consumption and related emissions, that would have occurred without the subject program. Baseline definitions consist of site-specific issues and broader, policy-oriented considerations.

Site-specific issues include the characteristics of equipment in place before an efficiency measure is implemented and how and when the affected equipment or systems are operated. For example, for an energy-efficient lighting retrofit, the baseline decisions include the type of lighting equipment that was replaced, the power consumption (watts/fixture) of the replaced equipment, and how many hours the lights would have operated. The broader baseline policy issues involve ensuring that the energy and demand savings and avoided emissions are "additional" to any that would otherwise occur due, for example, to federal or state energy standards.

When defining the baseline, it is also important to consider where in the life-cycle of the existing equipment or systems the new equipment was installed.

The options are (a) *early replacement* of equipment that had not reached the end of its useful life; (b) new, energy efficient equipment installed for failed *equipment replacement*; or (c) *new construction*. For each of these options, the two generic approaches to defining baselines are the *project-specific* and the *performance standard* procedure.

Project-Specific Baseline

Under the project-specific procedure (used on all or a sample of the projects in a program), the baseline is defined by a specific technology or practice that would have been pursued, at the site of individual projects, if the program had not been implemented. With energy efficiency programs, the common way this is accomplished is through (a) an assessment of the existing equipment's consumption rate, based on measurements or historical data; (b) an inventory of pre-retrofit equipment; or (c) a control group's energy equipment (used where no standard exists or when the project is an "early replacement"—that is, implemented prior to equipment failure).[2] Most organizations, when calculating their own savings, define baseline as what the new equipment actually replaces; that is, the baseline is related to actual historical base year energy consumption or demand. Note that because identifying this type of baseline always involves some uncertainty with respect to free riders, this approach should be used in combination with explicit additionality considerations.

Performance Standard Baseline

The second approach to determining baselines is to avoid project-specific determinations (and thus most free ridership issues) and instead try to ensure the additionality of quantified energy and demand savings, and avoided emissions. This is done by developing a performance standard, which provides an estimate of baseline energy and demand for all the projects in a program. The assumption is that any project activity will produce *additional* savings and avoided emissions if it has a "lower" baseline than the performance standard baseline. Performance standards are sometimes referred to as "multi-project baselines" because they can be used to estimate baseline emissions for multiple project activities of the same type.

Under the performance standard procedure, baseline energy and demand are estimated by calculating an average (or better-than-average) consumption rate (or efficiency) for a blend of alternative technologies or practices. These standards are used in large-scale retrofit (early replacement) programs when the range of equipment being replaced and how it is operated cannot be individually determined. This would be the case, for example, in a residential compact fluorescent incentive program, where the types of lamps being replaced and the number of hours they operate cannot be determined for each home. Instead, studies are used to determine typical conditions.

Another very common use of performance standards is to define a baseline as the minimum efficiency standard for a piece of equipment as defined by a law, code, or standard industry practice (often used for new construction or equipment that replaces failed equipment).

New Construction Baselines

It can be difficult to define baselines and additionality for new construction programs. This is somewhat apparent in that there are no existing systems to which the reporting period energy consumption and demand can be compared. However, the concepts of project and performance standard baseline definitions can still be used. The common ways in which new construction baselines are defined are:

- What would have been built or installed without the program at the specific site of each of project? This might be evaluated by standard practice or plans and specifications prepared prior to the program being introduced.
- Building codes and/or equipment standards.
- The performance of equipment, buildings, etc., in a comparison group of similar program non-participants.

Defining Adjustment Factors

As discussed in Chapter 4, the "adjustments" distinguish properly determined savings from a simple comparison of energy usage before and after implementation of a program. By accounting for factors (independent variables) that are beyond the control of the program implementer or energy consumer, the adjustments term brings energy use in the two time periods to the same set of conditions. Common examples of adjustment are:

- **Weather corrections**—for example, if the program involves heating or air-conditioning systems in buildings.
- **Occupancy levels and hours**—for example, if the program involves lighting retrofits in hotels or office buildings.
- **Production levels**—for example, if the program involves energy efficiency improvements in factories.

The choice of independent variables can be a major effort, as it involves testing which variables are influential. This is typically done during the implementation phase as part of data analysis efforts, but can also occur during the planning phase with significant variables identified on the basis of intuition and experience.

Defining Data Collection Requirements

Assessing baseline and adjustment issues in the planning stage is important for determining data collection and budgeting requirements. The goal is to avoid reaching the analysis stage of an evaluation and discovering that critical pieces of information have either not been collected or have been collected with an unreliable level of quality. These scenarios can be guarded against by providing specific instructions to program administrators and others. This may be necessary because the information needed for calculating benefits is not always useful to program administrators for their tasks of managing and tracking program progress. Planning for data collection is necessary to give administrators notice and justification for collecting items of data they would not ordinarily collect.

7.2.5 Establishing a Budget in the Context of Information Quality Goals

This subsection helps the evaluation planner define the accuracy expected for evaluation results. It also helps establish the overall evaluation budget, given the seven major issues identified in this chapter.

Establishing a budget (i.e., funding level) for an evaluation requires consideration of all aspects of the evaluation process, particularly the six other issues raised in this chapter. This subsection, however, discusses budgeting in the context of determining the appropriate level of certainty for the evaluation results.

California Example of Risk Management Approach to Evaluation Budgeting

California has a $170 million budget for evaluation studies and $70 million for impact studies. However, it still does not have enough money for rigorous evaluations of all but the most important and high-risk programs (i.e., those programs for which accuracy of findings is critical).

To help assign evaluation resources, California used a risk analysis approach that weighed the need for confidence and precision with the risk of the answer being wrong at the program level, the technology level, and the portfolio level. A prioritized list of programs was prepared in which the rigor levels of the evaluations could be structured to match the need for reliable information. The budget was then distributed to match the need. California used the Crystal Ball® analysis program. Using sets of possible error distributions (shapes) at the technology level for kWh, kW, and therms saved, a few hundred thousand risk analysis runs were made based on the probability of the assigned distribution shapes and the expected savings within those shapes.

Provided by Nick Hall of TecMarket Works.

When designing and implementing a program, the primary challenge associated with evaluation is typically *balancing (a) the cost, time and effort to plan and complete, and uncertainty of various approaches with (b) the value of the information generated by the efforts*. Most of the value of information is tied to the value of energy savings and overall program integrity. The costs for high levels of confidence in the calculations must be compared to the risks (and costs) associated with the value of savings being allocated to projects and programs. In this sense, evaluation processes are about risk management. Low-risk projects require less evaluation confidence and precision; high-risk projects require more confidence and precision. The acceptable level of uncertainty is often a subjective judgment based on the value of the energy and demand savings, the risk to the program associated with over- or underestimated savings, and a balance between encouraging efficiency actions and high levels of certainty. An important aspect of evaluation planning is deciding what level of risk is acceptable and thus the requirements for accuracy and a corresponding budget.

How much risk is acceptable is usually related to:

- The amount of savings expected from the program.
- Whether the program is expected to grow or shrink in the future.
- The uncertainty about expected savings and the risk the program poses in the context of achieving portfolio savings goals.
- The length of time since the last evaluation and the degree to which the program has changed in the interim.
- The requirements of the regulatory commission or oversight authority, and/or the requirements of the program administrator.

On a practical level, the evaluation budget reflects a number of factors. At the portfolio level, for example, evaluation budgets may be established in regulatory proceedings. At the program level, budgets are often influenced by factors that affect the level of quality associated with evaluation results. For example, budgets may increase to accomodate follow-up studies aimed at assessing and reducing measurement error, or to pay for additional short-term metering, training of staff, or testing of questionnaires and recording forms to reduce data collection errors. Additional resources might be required to ensure that "hard-to-reach" portions of the population are included in the sampling frame (reducing non-coverage error) or devoted to follow-up aimed at increasing the number of sample members for whom data are obtained (reducing non-response bias).

The determination of the appropriate sample size also affects the evaluation budget. To address this, procedures such as a statistical power analysis help researchers determine the sample size needed to achieve the desired level of precision and confidence for key outcomes so that those of a substantively important magnitude will be statistically significant. Appendix D discusses the steps that can be taken to increase the accuracy of evaluation results.

While it is difficult to generalize, the National Action Plan suggests that a reasonable spending range for evaluation is 3 to 6 percent of program budgets[3] In general, on a unit-of-saved-energy basis, costs are inversely proportional to the magnitude of the savings (i.e., larger projects have lower per-unit evaluation costs) and directly proportional to uncertainty of predicted savings (i.e., projects with greater uncertainty in the predicted savings warrant higher EM&V costs).

7.2.6 Selecting Impact Evaluation Approaches for Energy Savings and Avoided Emissions Calculations

This subsection reiterates the reasons for calculating gross or net energy savings and the various approaches for calculating net and gross energy savings and avoided emissions.

Chapters 4, 5, and 6 define approaches and present criteria for selecting approaches for determining gross and net energy and demand savings, as well as avoided emissions estimates. These will not be repeated here, but deciding (a) which of these results will be determined and (b) which of the calculation approaches will be used is a critical part of the planning process.

For completeness, the major calculation approaches are listed again below.

- For gross energy and demand savings, one or more of the following calculation approaches are used:
  - One or more M&V methods from the IPMVP are used to determine the savings from a sample of projects, and these savings are then applied to all of the projects in the program.
  - Deemed savings based on historical, verified data are applied to conventional energy efficiency measures implemented in the program.
  - Statistical analyses of large volumes of energy meter data are conducted.
- For net energy and demand savings, the calculation approaches are:
  - Self-reporting surveys.
  - Enhanced self-reporting surveys.
  - Econometric methods.
  - Stipulated NTGR.
- Related to the choice of net energy and demand savings approach are the factors used to convert gross to net savings. Thus, these should be selected concurrently. Factors for consideration are:
  - Free ridership.
  - Spillover.
  - Rebound.
  - T&D losses (electricity efficiency).
  - Economy factors and energy prices (or others).
- Avoided emission factor calculation approaches involve using:
  - System average emission rates.
  - Dispatch models.
  - Medium effort calculation approaches.

The decision to calculate net or gross energy savings depends on the program objectives and available evaluation resources. Gross savings are calculated when all that is needed is an estimate of the savings for each project participating in a program. The most common example of this is projects involving a contractor completing energy efficiency measures in facilities for the sole purpose of achieving energy savings (e.g., performance contracts). Net savings are calculated when it is of interest to know the level of savings that occurred as a result of the program's influence on program participants and non-participants. This is usually the case when public or ratepayer monies fund the evaluation program, and when accurate avoided emission estimates are desired.

7.2.7 Selecting An Evaluator

This subsection helps the evaluation planner select the evaluator.

Either the program implementer or a third party typically conducts the evaluation. The third party—valued for a more independent perspective—can be hired either by the implementer, with criteria for independence, or by an overseeing entity such as a utility regulator. A typical approach for utility-sponsored efficiency programs is for the utility's evaluation staff to manage studies that are completed by third-party consultants, whose work is reviewed by the utility regulatory agency. The objective is for all parties to the evaluation to believe that the reported results are based on valid information and are sufficiently reliable to serve as the basis for informed decisions.

There are advantages and disadvantages to using either implementers or independent third parties to conduct evaluations—selecting one or the other depends on the goals of the evaluation. Regulated energy programs and programs with a financial outcome hinging on the results of the evaluation tend to require third-party evaluation. Another approach is to have the evaluation completed by the implementer with the requirement for third-party verification. Some emission programs, such as the European Trading System for greenhouse gases, require third-party independent verification of avoided emissions information.

On the other hand, a common objective of evaluation is to improve the performance of a program and help with program improvement. This latter objective favors a tight relationship between the evaluator and the implementer. Thus, the selection of an evaluator can require balancing evaluation independence (so that the evaluation is objective) with the desire to have the evaluator close enough to the process such that the evaluation provides ongoing and early feedback without the implementer feeling "defensive."

Evaluators can either be in-house staff or consultants. Evaluation consulting firms tend to use either econometricians (professionals who apply statistical and mathematical techniques to problem solving) and engineers. Many are members of industry professional organizations, or are Certified Measurement and Verification Professionals (CMVPs).[4] Two of the professional organizations that energy evaluators participate in are:

- Association of Energy Service Professionals, <http://www.aesp.org>.
- International Energy Program Evaluation Conference, <http://www.iepec.org>.

In addition, the California Measurement Advisory Council (CALMAC) now offers a directory of evaluators: <http://www.calmac.org/contractorcontact.asp>.

7.3 Evaluation and M&V Plan Outlines

The program evaluation plan should be a formal document that clearly presents the evaluation efforts and details the activities to be undertaken during the evaluation. The evaluation plan is a stand-alone decision document, meaning it must contain the information the evaluator and others need to understand what is to be undertaken and how. The plan is also an important historical document in that it is not unusual for programs with long life cycles to undergo staff changes.

The following subsections outline the contents of an impact evaluation plan and an M&V plan. The M&V plan is included because it is a common approach for calculating gross energy savings. Following the M&V plan outline are evaluation planning checklists.

7.3.1 Evaluation Plan and Report Outlines

The following is a template that can be used to produce an impact evaluation plan.

A. Program Background

1. Short description of the program(s) being evaluated (e.g., the market, approach, technologies, budget, objectives, etc.).
2. Presentation of how the program will save energy and demand, and avoid emissions.
3. List of the technologies offered by the program.
4. Program schedule.
5. Numerical savings and avoided emission goals.

B. Evaluation Overview

1. List of evaluation objectives and how they support program goals.
2. List of which indicators will be reported (e.g., annual MWh, monthly peak kW, annual therms, annual CO_2).
3. Gross and net impact evaluation approaches selected and methodology for calculating avoided emissions, as appropriate.
4. List of primary factors will be considered in analysis of gross and net savings (e.g. weather, occupancy, free riders, spillover).
5. Budget and schedule summary.
6. Listing of evaluators (if known) or evaluator selection method.

C. Detailed Evaluation Approach, Scope, Budget, Schedule, and Staffing

This is the detailed presentation of the evaluation activities to be undertaken, including the M&V option to be used, as appropriate.

1. Gross impact savings analysis—a description of the analysis activities and approaches. (If an M&V evaluation approach is selected, identify the IPMVP Option to be used.)
2. Net impact savings analysis—a description of how spillover, free ridership, and other effects will be addressed in the evaluation activities and in the data analysis.
3. Data collection, handling, and sampling:
   - Measurement collection techniques.
   - Sampling approach and sample selection methods for each evaluation activity that includes sampling efforts.
   - How the comparison group, or non-participant, information will be used in the evaluation(s) and in the analysis.
   - Data handling and data analysis approach to be used to address the researchable issues.
4. Uncertainty of results—presentation and discussion of the threats to validity, potential biases, methods used to minimize bias, and level of precision and confidence associated with the sample selection methods and the evaluation approaches. Quality control information should also be included here.
5. An activities timeline with project deliverable dates.
6. Detailed budget.
7. Evaluation team—information concerning the independence of the evaluator. Evaluator contact information should be included here.

The final product or output of an evaluation is a report. The following is a sample report outline (taken from DOE, 2003):

- Table of Contents
- List of Figures and Tables
- Acronyms
- Abstract
- Acknowledgments

1. Executive Summary
   (Include highlights of key recommended improvements to the program, if relevant.)
2. Introduction
   - Program Overview (e.g., program description, objectives)
   - Evaluation Objectives and Methods
   - Structure of the Report
3. Study Methodology
   - Data Collection Approach(es)
   - Analysis Methods
   - Limitations, Caveats
4. Key Evaluation Results
   Answers for all of the questions specified for the evaluation. Could include several sections on findings. Findings could be presented for each method used, by program components covered, by market segments covered, and so forth, followed by a section on integrated findings or organized and presented by the different observed effects or type of results.
5. Recommendations
   If relevant; depends on the type of evaluation. Should include clear, actionable, and prioritized recommendations that are supported by the analysis.
6. Summary and Conclusions
7. Appendices (examples listed below):
   - Recommended improvements to the evaluation process, including any lessons learned for future evaluation studies.

- Appendices containing detailed documentation of the research design and assumptions, data collection methods, evaluation analysis methodology, results tables, etc.
- Survey or interview instrument, coding scheme, and compiled results tables and data.
- Sources and quality (caveats on data) of primary and secondary information.
- Details on quantitative data analysis: analytical framework, modeling approach, and statistical results.
- Qualifications and extensions.
- Possible sources of overestimation and underestimation.
- Treatment of issues concerning double counting, use of savings factors, synergistic effects.
- How attribution was addressed (for impact evaluation).
- Sensitivity of energy savings estimates.
- Assumptions and justifications.

7.3.2 M&V Plan Outline

If the M&V gross impact evaluation approach is selected, an M&V plan needs to be prepared that is applicable to each project selected for analysis. This section discusses the M&V planning process for individual projects and then presents an M&V plan outline.

M&V activities fall into five areas:

1. Selecting one of the four IPMVP Options for the project. The Options define general approaches to documenting savings.
2. Preparing a project-specific M&V plan that outlines the details of what will be done to document savings.
3. Defining the pre-installation baseline, including equipment and systems, baseline energy use, or factors that influence baseline energy use.
4. Defining the reporting period situation, including equipment and systems, post-installation energy use, and factors that influence post-installation

Evaluability

"Evaluability," a relatively new addition to the evaluation lexicon, is basically an assessment protocol to increase the probability that evaluation information will be available when evaluations are actually undertaken. Some data (for example, the age of a building) can be gathered at any time; some data are best gathered at the time of evaluation (participant spillover, current hours of operation); and some data must be gathered at the time of implementation or they will be lost forever or rendered unreliable due to changes in personnel or fading recollection (free ridership, removed equipment, or non-participant customer contact). The list below is an example of some of the items included in an evaluability assessment template:

- Is there a way to track participants?
- Is there a way to track non-participants?
- Are specific locations of measures being tracked? Can they be found?
- Are program assumptions being tracked on a site-specific level (e.g., hours of operation)?
- Is the delivered energy saving service and/or installed retrofit being recorded?
- Does the device recording savings include the outcome or result of the activities?
- Are savings assumptions documented?
- Is the source of savings assumptions specified?
- Are the pre-retrofit or baseline parameters being recorded?
- Does the database record the "as-found" values for parameters used to estimate ex ante savings?
- Does baseline monitoring need to take place?
- Can one of the impact evaluation methods specified in this Guide be used?
- Are there code compliance or program overlap issues for savings estimation?

energy use. Site surveys; spot, short-term, or long-term metering; and/or analysis of billing data can also be used for the reporting period assessment.

5. Conducting periodic (typically annual) M&V activities to verify the continued operation of the installed equipment or system, determine current year savings, identify factors that may adversely affect savings in the future, and estimate savings for subsequent years.

A project-specific M&V plan describes in reasonable detail what will be done to document the savings from a project. It can be a plan for each energy efficiency measure included in the project—for example, when a retrofit isolation approach is used. Or, it can cover the entire project—for example, when the whole-facility analyses approach is used. Regardless, the M&V plan will consider the type of energy efficiency measures involved and the desired level of accuracy.

The M&V plan should include a project description, facility equipment inventories, descriptions of the proposed measures, energy savings estimates, a budget, and proposed construction and M&V schedules. A project-specific M&V plan should demonstrate that any metering and analysis will be done consistently, logically, and with a level of accuracy acceptable to all parties.

The following is a sample M&V plan outline:

1. Description of project, measures to be installed, and project objectives.
2. Selected IPMVP Option and measurement boundary.
3. Description of base year conditions, data collection, and analyses.
4. Identification of any changes to base year conditions and how they will be accounted for.
5. Description of reporting period conditions, data collection, and analyses.
6. Basis for adjustments that may be made to any measurements and how this will be done.
7. Specification of exact analysis procedures.
8. Metering schedule and equipment specifications.

Table 7-1. Energy Efficiency Project M&V Plan Content—General Components

Category	M&V Plan Components
Project Description	Project goals and objectives
	Site characteristics and constraints (e.g., absence of utility meter data at site)
	Measure descriptions that include how savings will be achieved
Project Savings and Costs	Estimated savings by measure
	Estimated M&V cost by measure
Scheduling	Equipment installations
	M&V activities
Reporting	Raw data format
	Compiled data format
	Reporting interval
M&V Approach	Confidence and precision requirements
	Options used
	Person(s) responsible for M&V activities

Table 7-2. Energy Efficiency Project-Specific M&V Plan Contents—Measure-Specific Components

Category	M&V Plan Components	Examples
Analysis Method	Data requirements	kW, operating hours, temperature
	Basis of stipulated values	Lighting operating hours equal 4,000/year based on metered XYZ building
	Savings calculation equations	$\text{kWh savings} = [(\text{kW/Fixture}_{\text{baseline}} \times \text{Quantity}_{\text{baseline}}) - (\text{kW/Fixture}_{\text{post}} \times \text{Quantity}_{\text{post}})] \times \text{Operating Hours}$
	Regression expressions	Three parameter change-point cooling model
	Computer simulation models	DOE-2 simulation model
Metering and Monitoring	Metering protocols	ASHRAE Guideline 14 pump multiple point test throughout short-term monitoring
	Equipment	ABC Watt Hour Meter
	Equipment calibration protocols	National Institute of Science and Technology protocols
	Metering points	Flow rate, RMS power
	Sample size	25 lighting circuits out of 350
	Sampling accuracy	90% confidence/10% precision
	Metering duration and interval	2 weeks/15-minute data
Baseline Determination	Performance factors	Boiler efficiency
	Operating factors	Load, operating hours
	Existing service quality	Indoor temperature set points
	Minimum performance standards	State energy code
Savings Adjustments	Party responsible for developing adjustments	Smith Engineers, hired by sponsor
	Savings adjustment approach	Baseline adjusted for reported period weather and building occupancy levels

9. Description of expected accuracy and how it will be determined.

10. Description of quality assurance procedures.

11. Description of budget and schedule.

12. Description of who will conduct M&V.

The following tables summarize what could be contained in the M&V plans. Table 7-2 lists general requirements for an overall plan. Table 7-3 lists requirements that could be addressed for each measure (e.g., building lighting retrofit, building air conditioning retrofit, control system upgrade) that is included in the project being evaluated. More information on the contents of an M&V Plan can be found in the IPMVP (EVO, 2007).

7.3.3 Checklist of Planning Decisions for an Impact Evaluation

The following table presents a checklist for preparing an impact evaluation plan. The list is organized around the decisions associated with the gross savings calculation, net savings calculation, calculation of avoided emissions, and generic issues.

Table 7-3. Checklist of Planning Decisions for Impact Evaluation

Checklist for Gross Impact Evaluation (Chapter 4)

Savings to Be Reported	
Energy savings (annual, seasonal, monthly, hourly, other)	
Demand savings (peak, coincident, average, other)	
Selected Gross Energy Savings Calculation Approach	
Measurement and verification approach	
Deemed savings approach	
Large-scale billing analysis approach	
Quality assurance approach	
Measurement and Verification Approach	
IPMVP Option A, B, C, or D	
Deemed Savings Approach	
Source of deemed savings identified and verified	
Large-Scale Billing Analysis Approach	
Time-series comparison	
Control group comparison	
Control group, time-series comparison	
Sample Size Criteria Selected	

Checklist for Net Impact Evaluation (Chapter 5)

Net Savings Factors to Be Evaluated	
Free ridership	
Spillover effects	
Rebound effect	
Electricity T&D losses	
Other(s)	
Net Savings Calculation Approach Selected	
Self-reporting surveys	
Enhanced self-reporting surveys	
Econometric methods	
Stipulated net-to-gross ratio	
Sample Size Criteria Selected	

Checklist for Avoided Emissions Calculations (Chapter 6)

Electricity efficiency savings—grid-connected	
Operating or build margin evaluated, or both	
System average emissions rate	
Hourly dispatch model emissions rate	
Middle ground emissions rate	
Natural Gas, Fuel Oil, and Non-Grid-Connected Electric Generating Units	
Default emission factor	
Source testing	

Table 7-3 (continued). Checklist of Planning Decisions for Impact Evaluation

Generic Evaluation Considerations

Overall Goals	
Does the evaluation address the key policy, regulatory, and oversight needs for evaluation information?	
Will the program success in meeting energy, demand, and emissions goals be quantifiably evaluated in the same manner as they are defined for the program?	
Does the evaluation plan represent a reasonable approach to addressing the information needs?	
Are there missing opportunities associated with the evaluation approach that should be added or considered? Are any additional co-benefits being evaluated?	
Does the impact evaluation provide the data needed to inform other evaluations that may be performed, particularly cost-effectiveness analyses?	
Has a balance been reached between evaluation costs, uncertainty of results, and value of evaluation results?	
Uncertainty of Evaluation Results	
Can the confidence and precision of the evaluation results be quantified? If so, how?	
Are there key threats to the validity of the conclusions? Are they being minimized given budget constraints and study tradeoffs? Will they be documented and analyzed?	
Is the evaluation capable of providing reliable conclusions on energy and other impacts?	
Budget, Timing, and Resources	
Does the evaluation take advantage of previous evaluations and/or concurrent ones for other programs?	
Does the cost of the study match the methods and approaches planned?	
Do the scheduled start and end times of the evaluation match the need for adequate data gathering, analysis, and reporting?	
Are adequate human resources identified?	
Does the evaluation rely on data and project access that are reasonably available?	
Reporting	
Are the time frames and scopes of evaluation reported defined?	
Do the data collection, analysis, and quality control match the reporting needs?	
Are the persistence of savings and avoided emissions being evaluated?	
Have measurement and impacts (emissions) boundaries been properly set?	
Sampling and Accuracy	
Is the sampling plan representative of the population served?	
Is the sampling plan able to support the evaluation policy objectives?	
Are there threats to the validity of the evaluation results that are incorporated into the evaluation design?	

7.4 Notes

1. A companion National Action Plan document that addresses program planning is the *Guide to Resource Planning with Energy Efficiency*, available at <www.epa.gov/eeactionplan>.

2. In early replacement projects, a consideration in whether to use existing conditions or code requirements for a baseline is if the replaced equipment or systems had a remaining lifetime shorter that the time period of the evaluation. In this situation, the first year(s) of the evaluation might have an existing condition baseline and the later years a code requirements baseline.

3. Robust evalution budgets may be supported when:

   - Energy efficiency is being used as an alternative to new supply. If the savings from energy efficiency programs is a significant factor in resource planning (e.g., energy efficiency is being considered an alternative to new supply), then it is particularly important that evaluation results be robust.

   - Many programs are being run. In some states, many different organizations administer energy efficiency programs. This results in a proliferation of programs to evaluate and a resulting increase in evaluation budgets. In addition, some states with fewer program administrators run many different programs, or variations on programs, for different target audiences.

   - A shareholder incentive or other program administrator reward program exists. Awarding incentives for meeting energy efficiency goals can increase the need for more robust evaluation results in order to reduce the possibility of contention over the results and associated awards.

4. The CMVP program is a joint activity of the Efficiency Valuation Organization and the Association of Energy Engineers (AEE). It is accessible through EVO’s Web site, <http://www.evo-world.org>.

Appendix A: National Action Plan for Energy Efficiency Leadership Group

Co-Chairs

Marsha Smith
Commissioner, Idaho Public Utilities Commission
President, National Association of Regulatory Utility Commissioners

James E. Rogers
Chairman, President, and C.E.O.
Duke Energy

Leadership Group

Barry Abramson
Senior Vice President
Servidyne Systems, LLC

Tracy Babbidge
Director, Air Planning
Connecticut Department of Environmental Protection

Angela S. Beehler
Director of Energy Regulation
Wal-Mart Stores, Inc.

Jeff Bladen
General Manager, Market Strategy
PJM Interconnection

Sheila Boeckman
Manager of Business Operations and Development
Waverly Light and Power

Bruce Braine
Vice President, Strategic Policy Analysis
American Electric Power

Cheryl Buley
Commissioner
New York State Public Service Commission

Jeff Burks
Director of Environmental Sustainability
PNM Resources

Kateri Callahan
President
Alliance to Save Energy

Jorge Carrasco
Superintendent
Seattle City Light

Lonnie Carter
President and C.E.O.
Santee Cooper

Gary Connett
Manager of Resource Planning and Member Services
Great River Energy

Larry Downes
Chairman and C.E.O.
New Jersey Natural Gas (New Jersey Resources Corporation)

Roger Duncan
Deputy General Manager, Distributed Energy Services
Austin Energy

Angelo Esposito
Senior Vice President, Energy Services and Technology
New York Power Authority

Jeanne Fox
President
New Jersey Board of Public Utilities

Anne George
Commissioner
Connecticut Department of Public Utility Control

Dian Grueneich
Commissioner
California Public Utilities Commission

Blair Hamilton
Policy Director
Vermont Energy Investment Corporation

Leonard Haynes
Executive Vice President, Supply Technologies, Renewables, and Demand Side Planning
Southern Company

Mary Healey
Consumer Counsel for the State of Connecticut
Connecticut Consumer Counsel

Joe Hoagland
Vice President, Energy Efficiency and Demand Response
Tennessee Valley Authority

Sandy Hochstetter
Vice President, Strategic Affairs
Arkansas Electric Cooperative Corporation

Helen Howes
Vice President, Environment, Health and Safety
Exelon

Bruce Johnson
Director, Energy Management
Keyspan

Mary Kenkel
Consultant, Alliance One
Duke Energy

Ruth Kiselewich
Director, Conservation Programs
Baltimore Gas and Electric

Rick Leuthauser
Manager of Energy Efficiency
MidAmerican Energy Company

Harris McDowell
Senator
Delaware General Assembly

Mark McGahey
Manager
Tristate Generation and Transmission Association, Inc.

Ed Melendreras
Vice President, Sales and Marketing
Entergy Corporation

Janine Migden-Ostrander
Consumers' Counsel
Office of the Ohio Consumers' Counsel

Michael Moehn
Vice President, Corporate Planning
Ameren Services

Fred Moore
Director, Manufacturing and Technology, Energy
The Dow Chemical Company

Richard Morgan
Commissioner
District of Columbia Public Service Commission

Brock Nicholson
Deputy Director
Division of Air Quality
North Carolina Air Office

Pat Oshie
Commissioner
Washington Utilities and Transportation Commission

Douglas Petitt
Vice President,
Government Affairs
Vectren Corporation

Bill Prindle
Deputy Director
American Council for an Energy-Efficient Economy

Phyllis Reha
Commissioner
Minnesota Public Utilities Commission

Roland Risser
Director, Customer Energy Efficiency
Pacific Gas and Electric

Gene Rodrigues
Director, Energy Efficiency
Southern California Edison

Art Rosenfeld
Commissioner
California Energy Commission

Gina Rye
Energy Manager
Food Lion

Jan Schori
General Manager
Sacramento Municipal Utility District

Ted Schultz
Vice President,
Energy Efficiency
Duke Energy

Larry Shirley
Division Director
North Carolina Energy Office

Tim Stout
Vice President, Energy Efficiency
National Grid

Deb Sundin
Director, Business Product Marketing
Xcel Energy

Paul Suskie
Chairman
Arkansas Public Service Commission

Dub Taylor
Director
Texas State Energy Conservation Office

Paul von Paumgartten
Director, Energy and Environmental Affairs
Johnson Controls

Brenna Walraven
Executive Director, National Property Management
USAA Realty Company

Devra Wang
Director, California Energy Program
Natural Resources Defense Council

J. Mack Wathen
Vice President, Regulatory Affairs
Pepco Holdings, Inc.

Mike Weedall
Vice President, Energy Efficiency
Bonneville Power Administration

Zac Yanez
Program Manager
Puget Sound

Henry Yoshimura
Manager, Demand Response
ISO New England Inc.

Dan Zaweski
Assistant Vice President of Energy Efficiency and Distributed Generation
Long Island Power Authority

Observers

Keith Bissell
Attorney
Gas Technology Institute

Rex Boynton
President
North American Technician Excellence

James W. (Jay) Brew
Counsel
Steel Manufacturers Association

Roger Cooper
Executive Vice President,
Policy and Planning
American Gas Association

Dan Delurey
Executive Director
Demand Response Coordinating Committee

Reid Detchon
Executive Director
Energy Future Coalition

Roger Fragua
Deputy Director
Council of Energy Resource Tribes

Jeff Genzer
General Counsel
National Association of State Energy Officials

Donald Gilligan
President
National Association of Energy Service Companies

Chuck Gray
Executive Director
National Association of Regulatory Utility Commissioners

Steve Hauser
President
GridWise Alliance

William Hederman
Member, IEEE-USA Energy Policy Committee
Institute of Electrical and Electronics Engineers

Marc Hoffman
Executive Director
Consortium for Energy Efficiency

John Holt
Senior Manager of Generation and Fuel
National Rural Electric Cooperative Association

Eric Hsieh
Manager of Government Relations
National Electrical Manufacturers Association

Lisa Jacobson
Executive Director
Business Council for Sustainable Energy

Kate Marks
Energy Program Manager
National Conference of State Legislatures

Joseph Mattingly
Vice President, Secretary and General Counsel
Gas Appliance Manufacturers Association

Kenneth Mentzer
President and C.E.O.
North American Insulation Manufacturers Association

Diane Munns
Executive Director, Retail Energy
Edison Electric Institute

Michelle New
Director, Grants and Research
National Association of State Energy Officials

Ellen Petrill
Director, Public/Private Partnerships
Electric Power Research Institute

Alan Richardson
President and C.E.O.
American Public Power Association

Andrew Spahn
Executive Director
National Council on Electricity Policy

Rick Tempchin
Director, Retail Distribution Policy
Edison Electric Institute

Mark Wolfe
Executive Director
Energy Programs Consortium

Facilitators

U.S. Department of Energy

U.S. Environmental Protection Agency

Appendix B: Glossary

This glossary is based primarily on three evaluation-related reference documents:

1. 2007 IPMVP
2. 2004 California Evaluation Framework
3. 2006 DOE EERE Guide for Managing General Program Evaluation Studies

In some cases, the definitions presented here differ slightly from the reference documents. This is due to discrepancies across documents and author interpretations.

Additionality: A criterion that says avoided emissions should only be recognized for project activities or programs that would not have "happened anyway." While there is general agreement that additionality is important, its meaning and application remain open to interpretation.

Adjustments: For M&V analyses, factors that modify baseline energy or demand values to account for independent variable values (conditions) in the reporting period.

Allowances: Allowances represent the amount of a pollutant that a source is permitted to emit during a specified time in the future under a cap and trade program.. Allowances are often confused with credits earned in the context of project-based or offset programs, in which sources trade with other facilities to attain compliance with a conventional regulatory requirement. Cap and trade program basics are discussed at the following EPA Web site: <http://www.epa.gov/airmarkets/cap-trade/index.html>.

Analysis of covariance (ANCOVA) model. A type of regression model also referred to as a "fixed effects" model.

Assessment boundary: The boundary within which all the primary effects and significant secondary effects associated with a project are evaluated.

Baseline: Conditions, including energy consumption and related emissions, that would have occurred without implementation of the subject project or program. Baseline conditions are sometimes referred to as "business-as-usual" conditions. Baselines are defined as either project-specific baselines or performance standard baselines.

Baseline period: The period of time selected as representative of facility operations before the energy efficiency activity takes place.

Bias: The extent to which a measurement or a sampling or analytic method systematically underestimates or overestimates a value.

California Measurement Advisory Council (CALMAC): An informal committee made up of representatives of the California utilities, state agencies, and other interested parties. 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.

Co-benefits: The impacts of an energy efficiency program other than energy and demand savings.

Coincident demand: The metered demand of a device, circuit, or building that occurs at the same time as the peak demand of a utility's system load or at the same time as some other peak of interest, such as building or facility peak demand. This should be expressed so as to indicate the peak of interest (e.g., "demand coincident with the utility system peak") Diversity factor is defined as the ratio of the sum of the demands of a group of users to their coincident maximum demand. Therefore, diversity factors are always equal to one or greater.

Comparison group: A group of consumers who did not participate in the evaluated program during the program year and who share as many characteristics as possible with the participant group.

Conditional Savings Analysis (CSA): A type of analysis in which change in consumption modeled using regression analysis against presence or absence of energy efficiency measures.

Confidence: An indication of how close a value is to the true value of the quantity in question. Confidence is the likelihood that the evaluation has captured the true impacts of the program within a certain range of values (i.e., precision).

Cost-effectiveness: An indicator of the relative performance or economic attractiveness of any energy efficiency investment or practice. In the energy efficiency field, the present value of the estimated benefits produced by an energy efficiency program is compared to the estimated total costs 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 from a societal perspective).

Database for Energy-Efficient Resources (DEER): A California database designed to provide well-documented estimates of energy and peak demand savings values, measure costs, and effective useful life.

Deemed savings: An estimate of an energy savings or energy-demand savings outcome (gross savings) for a single unit of an installed energy efficiency measure that (a) has been developed from data sources and analytical methods that are widely considered acceptable for the measure and purpose and (b) is applicable to the situation being evaluated.

Demand: The time rate of energy flow. Demand usually refers to electric power measured in kW (equals kWh/h) but can also refer to natural gas, usually as Btu/hr, kBtu/hr, therms/day, etc.

Direct emissions: Direct emissions are changes in emissions at the site (controlled by the project sponsor or owner) where the project takes place. Direct emissions are the source of avoided emissions for thermal energy efficiency measures (e.g., avoided emissions from burning natural gas in a water heater).

Effective useful life: An estimate of the median number of years that the efficiency measures installed under a program are still in place and operable.

Energy efficiency: The use of less energy to provide the same or an improved level of service to the energy consumer in an economically efficient way; or using less energy to perform the same function. "Energy conservation" is a term that has also been used, but it has the connotation of doing without a service in order to save energy rather than using less energy to perform the same function.

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.

Engineering model: Engineering equations used to calculate energy usage and savings. These models are usually based on a quantitative description of physical processes that transform delivered energy into useful work such as heat, lighting, or motor drive. In practice, these models may be reduced to simple equations in spreadsheets that calculate energy usage or savings as a function of measurable attributes of customers, facilities, or equipment (e.g., lighting use = watts × hours of use).

Error: Deviation of measurements from the true value.

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, 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: Forecasted savings used for program and portfolio planning purposes. (From the Latin for "beforehand.")

Ex post evaluation estimated savings: Savings estimates reported by an evaluator after the energy impact evaluation has been completed. (From the Latin for "from something done afterward.")

Free driver: A non-participant who has adopted a particular efficiency measure or practice as a result of the evaluated program.

Free rider: A program participant who would have implemented the program measure or practice in the absence of the program. Free riders can be total, partial, or deferred.

Gross savings: The change in energy consumption and/or demand that results directly from program-related actions taken by participants in an efficiency program, regardless of why they participated.

Impact evaluation: An evaluation of the program-specific, directly induced changes (e.g., energy and/or demand usage) attributable to an energy efficiency program.

Independent variables: The factors that affect energy use and demand, but cannot be controlled (e.g., weather or occupancy).

Indirect emissions: Changes in emissions that occur at the emissions source (e.g., the power plant). Indirect emissions are the source of avoided emissions for electric energy efficiency measures.

Interactive factors: Applicable to IPMVP Options A and B; changes in energy use or demand occurring beyond the measurement boundary of the M&V analysis.

Leakage: In the context of avoided emissions, emissions changes resulting from a project or program not captured by the primary effect (typically the small, unintended emissions consequences). Sometimes used interchangeably with "secondary effects," although leakage is a more "global" issue whereas secondary, interactive effects tend to be considered within the facility where a project takes place.

Load shapes: Representations such as graphs, tables, and databases that describe energy consumption rates as a function of another variable such as time or outdoor air temperature.

Market effect evaluation: An evaluation of the change in the structure or functioning of a market, or the behavior of participants in a market, that results from one or more program efforts. Typically the resultant market or behavior change leads to an increase in the adoption of energy-efficient products, services, or practices.

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.

Measurement: A procedure for assigning a number to an observed object or event.

Measurement and verification (M&V): Data collection, monitoring, and analysis associated with the calculation of gross energy and demand savings from individual sites or projects. M&V can be a subset of program impact evaluation.

Measurement boundary: The boundary of the analysis for determining direct energy and/or demand savings.

Metering: The collection of energy consumption data over time through the use of meters. These meters may collect information with respect to an end-use, a circuit, a piece of equipment, or a whole building (or facility). Short-term metering generally refers to data collection for no more than a few weeks. End-use metering refers specifically to separate data collection for one or more end-uses in a facility, such as lighting, air conditioning or refrigeration. Spot metering is an instantaneous measurement (rather than over time) to determine an energy consumption rate.

Monitoring: Gathering of relevant measurement data, including but not limited to energy consumption 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).

Net savings: The total change in load that is attributable to an energy efficiency program. This change in load may include, implicitly or explicitly, the effects of free drivers, free riders, energy efficiency standards, changes in the level of energy service, and other causes of changes in energy consumption or demand.

Net-to-gross ratio (NTGR): A factor representing net program savings divided by gross program savings that is applied to gross program impacts to convert them into net program load impacts.

Non-participant: Any consumer who was eligible but did not participate in the subject efficiency program, in a given program year. Each evaluation plan should provide a definition of a non-participant as it applies to a specific evaluation.

Normalized annual consumption (NAC) analysis: A regression-based method that analyzes monthly energy consumption data.

Participant: A consumer that received a service offered through the subject efficiency 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 define "participant" as it applies to the specific evaluation.

Peak demand: The maximum level of metered demand during a specified period, such as a billing month or a peak demand period.

Persistence study: A study to assess changes in program impacts over time (including retention and degradation).

Portfolio: Either (a) a collection of similar programs addressing the same market (e.g., a portfolio of residential programs), technology (e.g., motor efficiency programs), or mechanisms (e.g., loan programs) or (b) the set of all programs conducted by one organization, such as a utility (and which could include programs that cover multiple markets, technologies, etc.).

Potential studies: Studies conducted to assess market baselines and savings potentials for different technologies and customer markets. Potential is typically defined in terms of technical potential, market potential, and economic potential.

Precision: The indication of the closeness of agreement among repeated measurements of the same physical quantity.

Primary effects: Effects that the project or program are intended to achieve. For efficiency programs, this is primarily a reduction in energy use per unit of output.

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 recommending improvements to increase the program's efficiency or effectiveness for acquiring energy resources while maintaining high levels of participant satisfaction.

Program: A group of projects, with similar characteristics and installed in similar applications. Examples could include a utility program to install energy-efficient lighting in commercial buildings, a developer's program to build a subdivision of homes that have photovoltaic systems, or a state residential energy efficiency code program.

Project: An activity or course of action involving one or multiple energy efficiency measures, at a single facility or site.

Rebound effect: A change in energy-using behavior that yields an increased level of service and occurs as a result of taking an energy efficiency action.

Regression analysis: Analysis of the relationship between a dependent variable (response variable) to specified independent variables (explanatory variables). The mathematical model of their relationship is the regression equation.

Reliability: Refers to the likelihood that the observations can be replicated.

Reporting period: The time following implementation of an energy efficiency activity during which savings are to be determined.

Resource acquisition program: Programs designed to directly achieve energy and or demand savings, and possibly avoided emissions

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 confidence and precision. The higher the level of rigor, the more confident one is that the results of the evaluation are both accurate and precise.

Secondary effects: Unintended impacts of the project or program such as rebound effect (e.g., increasing energy use as it becomes more efficient and less costly to use), activity shifting (e.g., when generation resources move to another location), and market leakage (e.g., emission changes due to changes in supply or demand of commercial markets). These secondary effects can be positive or negative.

Spillover: Reductions in energy consumption and/or demand caused by the presence of the energy efficiency program, beyond the program-related gross savings of the participants. There can be participant and/or non-participant spillover.

Statistically adjusted engineering (SAE) models: A category of statistical analysis models that incorporate the engineering estimate of savings as a dependent variable.

Stipulated values: See “deemed savings.”

Takeback effect: See “rebound effect.”

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.

Appendix
C: Other Evaluation Types

This appendix provides a brief introduction to additional types of evaluations, including process and market effects evaluations, cost-effectiveness analysis, and evaluations for three common program types (market transformation, codes and standards, and education and training). It is intended to supplement the body of the Guide, which is primarily focused on impact evaluations.

C.1 Process, Market Effects, and Cost-Effectiveness Evaluations

The following subsections briefly introduce two other, non-impact types of evaluations and cost-effectiveness analysis. These types of evaluations can involve interrelated activities and have interrelated results, and are often conducted at the same time. Table C-1 compares these three types plus impact evaluations.

C.1.1 Process Evaluations

The goal of process evaluations is to produce improved and more cost-effective programs. Thus, process evaluations examine the efficiency and effectiveness of program implementation procedures and systems. These evaluations usually consist of asking questions of those involved in the program, analyzing their answers, and comparing results to established best practices.

Process evaluations are particularly valuable when:

- The program is new or has many changes.
- Benefits are being achieved more slowly than expected.
- There is limited program participation or stakeholders are slow to begin participating.
- The program has a slow startup.
- Participants are reporting problems.
- The program appears not to be cost-effective.

Typical process evaluation results involve recommendations for changing a program's structure, implementation approaches, or program design, delivery, and goals.

The primary mechanism of process evaluations is data collection (e.g., surveys, questionnaires, and interviews) from administrators, designers, participants (such as

Table C-1. Program Evaluation Types

Evaluation Type	Description	Uses
Impact Evaluation	Quantifies direct and indirect benefits of the program.	Determines the amount of energy and demand saved, the quantity of emissions reductions, and possibly the co-benefits.
Process Evaluations	Indicates how the program implementation procedures are performing from both administration and participant perspectives.	Identifies how program processes can be improved.
Market Effects Evaluation	Indicates how the overall supply chain and market have been affected by the program.	Determines changes that have occurred in markets and whether they are sustainable with or without the program.
Cost-Effectiveness Evaluation	Quantifies the cost of program implementation and compares with program benefits.	Determines whether the energy efficiency program is a cost-effective investment as compared to other programs and energy supply resources.

Table C-2. Elements of a Typical Process Evaluation

• Program Design – The program mission – Assessment of program logic – Use of new practices or best practices	• Program Administration – Program oversight – Program staffing – Management and staff training – Program information and reporting
• Program Implementation – Quality control – Operational practice—how program is implemented – Program targeting, marketing, and outreach efforts – Program timing	• Participant Response – Participant interaction and satisfaction – Market and government allies interaction and satisfaction

facility operators), implementation staff (including contractors, subcontractors, and field staff), and key policy makers. Other elements of a process evaluation can include workflow and productivity measurements; reviews, assessments, and testing of records, databases, program-related materials, and tools; and possibly collection and analysis of relevant data from third-party sources (e.g., equipment vendors, trade allies).

Table C-2 lists examples of the issues that are typically assessed during a process evaluation.

C.1.2 Market Effects Evaluations

Program-induced changes that affect non-participants or the way a market operates are addressed in market effects evaluations. One way to think of these is that they estimate the effect a program has on future energy efficiency activities.

Market effects evaluations often involve a significant undertaking, since they are designed to determine whether the market is changing. For example, a market effects study could evaluate increases in the adoption of the products or services being promoted by the program (or more likely, a portfolio of programs). It might answer the question: Are vendors stocking and promoting more energy efficiency technologies as a result of the program? Market effects are sometimes called the ultimate test of a program's success, answering the question—will efficiency best practices continue in the marketplace, even after the current program ends?

Potential Studies

Another form of market study is called a potential study. Potential studies are conducted before a program is implemented in order to assess market baselines and savings potentials for different technologies and customer markets. These studies can also assess customer needs and barriers to adoption of energy efficiency, as well as how best to address these barriers through program design. Potential studies indicate what can be expected in terms of savings from a program. Potential is often defined in terms of *technical potential* (what is technically feasible given commercially available products and services), *economic potential* (which is the level of savings that can be achieved assuming a certain level of participant and/or societal cost-effectiveness is required), and *market potential* (what the market can provide, which is almost always less than market potential). Findings also help managers identify the program's key markets and clients and how to best serve the intended customers.

Market effects evaluations usually consist of surveys, reviews of market data, and analysis of the survey results and collected data. Some possible results from a market assessment include:

- Total market effects.
- An estimate of how much of the market effect is due to the program being evaluated.
- An estimate of whether the market effect is sustainable.

A market effects evaluation analyzes:

- Are the entities that undertook efficiency projects undertaking additional projects or incorporating additional technologies in their facilities that were not directly induced by the program? This might indicate that the facility operators have become convinced of the value of, for example, high-efficiency motors, and are installing them on their own.
- Are entities that did not undertake projects now adopting concepts and technologies that were encouraged by the program? This might indicate that the program convinced other facility operators of the advantages of the efficiency concepts.
- Are manufacturers, distributors, vendors, and others involved in the supply chain of efficiency products (and services) changing their product offerings, how they are marketing them, how they are pricing them, stocking them, etc.? The answers can indicate how the supply chain is adapting to changes in supply of and demand for efficiency products.

As can be deduced, the market effects evaluation can easily overlap with the spillover analyses conducted as part of an impact evaluation. Market effects studies, however, are interested in long-term, sustained effects, versus a more short-term spillover perspective. According to a study by the New England Efficiency Partnership (NEEP, 2006), most programs use direct participation and spillover as the basis for estimating market transformation program benefits, rather than projections of baselines and market penetration. Anecdotal evidence suggests that measurement of participant spillover is relatively common, while measurement of non-participant spillover is inconsistent across program administrators. About one fourth of the states in the 2006 study estimated ultimate effects by projecting change in market penetration relative to a projected baseline for at least some of their market transformation programs.

C.1.3 Cost-Effectiveness Analyses

Cost-effectiveness (sometime called cost-benefit) evaluations compare program benefits and costs, showing the relationship between the value of a program's outcomes and the costs incurred to achieve those benefits. The findings help program managers judge whether to retain, revise, or eliminate program elements and provide feedback on whether efficiency is a wise investment as compared to energy generation and/or procurement options. It also often a key component of the evaluation process for programs using public or utility ratepayer funds.

A variety of frameworks have historically been used to assess cost-effectiveness of energy efficiency initiatives. In the late 1970s, CPUC implemented a least-cost planning strategy in which demand-side reductions in energy use were compared to supply additions. One result of this strategy was *The Standard Practice Manual* (SPM). This document provided several methodologies for conducting cost-benefit analyses of utility-administered efficiency programs. The first version of the SPM was published in 1983. The document has been updated from time to time, with the most reason version dated 2001 (California State Governor's Office, 2001). The SPM is perhaps the definitive resource for information on cost-effectiveness tests for efficiency programs.

The SPM established several tests that can be used to evaluate the cost-effectiveness of publicly funded energy efficiency initiatives. These include the ratepayer impact measure test, the utility cost test, the participant test, the total resource cost test, and the societal test. These metrics vary in terms of (a) their applicability to different program types, (b) the cost and benefit elements included in the calculation, (c) the methods by which the cost and benefit elements are computed, and (d) the uses of the results. Most regulated utility efficiency programs use one or more versions of these tests, sometimes with variations unique to the requirements of a particular regulatory commission. Definitions of these tests (paraphrased from the SPM) are provided below.

- **Total resource cost (TRC) test.** The TRC test measures the net costs of a demand-side management program as a resource option based on the total costs of the program, including both the participants' and the utility's costs. The TRC ratio equals the benefits of the program, in terms of value of energy and demand saved, divided by the net costs. The ratio is usually calculated on a life-cycle basis considering savings and costs that accrue over the lifetime of installed energy efficiency equipment, systems, etc. When the TRC test is used, if the ratio is greater than 1.0, then the program is considered cost-effective, with of course proper consideration of uncertainties in the TRC ratio calculation. This is probably the most commonly applied cost-effectiveness test.
- **Utility cost (UC) test.** The UC test measures the net costs of a demand-side management program as a resource option based on the costs incurred by the administrator of the program (assumed to be a utility, though it can be any organization), excluding any net costs incurred by the participant. The benefits are the same as the TRC benefits (energy and demand savings value), but the costs are defined more narrowly and do not include consumer costs.
- **Participant test.** The participant test assesses cost-effectiveness from the participating consumer's perspective by calculating the quantifiable benefits and costs to the consumer of participating in a program. Since many consumers do not base their decision to participate entirely on quantifiable variables, this test is not necessarily a complete measure of all the benefits and costs a participant perceives.
- **Societal test.** The societal test, a modified version of the TRC, adopts a societal rather than a utility service area perspective. The primary difference between the societal and TRC tests is that, to calculate life cycle costs and benefits, the societal test accounts for externalities (e.g., environmental benefits), excludes tax credit benefits, and uses a societal discount rate.
- **Ratepayer impact measure (RIM) test.** The RIM test only applies to utility programs. It measures what happens to consumer bills or rates due to changes in utility revenues and operating costs caused by the program. This test indicates the direction and magnitude of the expected change in customer bills or rate levels.

C.2 Evaluating Other Program Types

This Guide focuses on the evaluation of programs whose primary goal is to directly achieve energy and demand savings and perhaps avoided emissions—resource acquisition programs. While all efficiency programs hope to achieve savings, some are designed to achieve these savings more indirectly. Evaluation of three other common program types (market transformation, codes and standards, and education and training) is briefly discussed below.

C.2.1 Market Transformation Programs

Market transformation (MT) denotes a permanent, or at least long-term, change in the operation of the market for energy efficiency products and services. MT programs attempt to reduce market barriers through market interventions that result in documented market effects that lasts after the program (intervention) has been withdrawn reduced or changed. During the 1990s, the focus of many energy efficiency efforts shifted from resource acquisition to market transformation. Subsequently there has been a shift back; resource acquisition, MT, and other program types are now implemented, often in a complementary manner. To a large extent, all programs can be considered MT in that they involve changing how energy efficiency activities take place in the marketplace.

MT evaluation tends to be a combination of impact, process, and market effect evaluation and can also include cost-effectiveness evaluations. However, given that the ultimate aim of MT programs is to increase the adoption of energy efficient technologies and practices, MT evaluation usually focuses first on energy efficiency adoption rates by market actors and second on the directly associated energy and demand savings. Also, MT programs are dynamic, and thus the nature of market effects can be expected to vary over time. Market actors that influence end-use consumer choices include installation and repair contractors, retailer staffs, architects, design engineers, equipment distributors, manufacturers, and of course the consumers themselves.

Evaluation plays an important role in providing the kind of feedback that can be used to refine the design of market interventions. This role is equally important for resource acquisition and MT interventions, but arguably more complex for MT programs since the interest is long-term changes in the market versus more immediate and direct energy savings for resource acquisition programs. Most importantly, evaluation for MT entails the collection of information that can be used to refine the underlying program theory (see side bar).

Evaluation of MT interventions also needs to focus on the mechanism through which changes in adoptions and energy usage are ultimately induced. This means that considerable attention must be focused on indicators of market effects through market tracking. Thus, a MT evaluation might first report changes in sales patterns and volumes for particular efficiency products as an indication of program progress in meeting program goals. (For more information on MT evaluation, see DOE, 2007).

C.2.2 Codes and Standards Programs

Most codes and standards programs involve (a) new or changed building codes or appliance and equipment standards and/or (b) increasing the level of compliance with code requirements or appliance standards. These programs are intended to save energy and demand and achieve co-benefits, primarily in new construction or major retrofits (for building codes) or when new equipment is purchased (appliance and equipment standards).

The primary approach to establishing energy and demand savings (and avoided emissions) values for the codes and standards programs is to assess the energy and demand impacts of the market adoption and decision changes caused by the new, modified, or better-enforced codes or standards and then adjust those savings

Theory-Based Evaluation: A Guiding Principle for MT Evaluation

Theory-based evaluation (TBE), an evaluation approach that has been widely used in the evaluation of social programs in other fields, has gained some foothold in the energy efficiency industry over the past few years. It involves a relatively detailed and articulated *program theory*, established up front, that specifies the sequence of events a program is intended to cause, along with the precise causal mechanisms leading to these events. Evaluation then focuses on testing the consistency of observed events with the overall program theory.

A TBE can be considered a process of determining whether a program theory is correct or not (i.e., testing a hypothesis). For example, with an incentive program, the theory is that paying a certain level of incentives will result in a certain level of energy and demand savings.

Having well-defined program theories helps focus an evaluation objective on assessing the validity of those theories, primarily to see whether a program concept is successful and should be expanded and/or repeated.

In the energy efficiency field to date, TBE is particularly well adapted to evaluating the effectiveness of market transformation initiatives. This is largely because market transformation tends to take a relatively long time to occur, involve a relatively large number of causal steps and mechanisms, and encompass changing the behavior of multiple categories of market actors, all of which makes it particularly fruitful to focus on specifying and testing a detailed and articulated program theory.

Provided by Ralph Prahl.

to account for what would have occurred if the code or standard change or enforcement did not occur. 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. For example, analysis of a new appliance standard would involve (a) estimating the life-cycle savings associated with each new appliance placed into service as compared to a standard practice or old-standard appliances, (b) multiplying those savings by the rate over time that the new appliances are placed into service, and (c) adjusting the resulting savings estimate by the number of high-efficiency appliances that consumers would have purchased even if the standard were not in place.

C.2.3 Education and Training Programs

Education and training programs only indirectly result in energy and demand savings. They can include advertising, public service announcements, education efforts, training activities, outreach efforts, demonstration projects, and other information- or communication-based efforts. These programs may be targeted to either end-use customers or other market actors whose activities influence the energy-related choices of end-use customers.

Typically, information and education programs have one or more of the following general goals:

- Educate energy consumers regarding ways to increase the energy efficiency of their facilities and activities, and thus convince them to take actions that help them manage their consumption or adopt more energy-efficient practices.
- Inform energy consumers and/or other market actors about program participation opportunities in order to increase enrollment in these programs.
- Inform energy consumers and/or other market actors about energy issues, behaviors, or products in an effort to transform the normal operations of the market.

Almost every energy efficiency program provides some level of educational and/or informational content. However, education-specific programs are typically designed to achieve energy or demand savings indirectly through changes in behavior, over time (market transformation) or via increased enrollments in other resource acquisition programs.

Understanding and Affecting Behavior

Some recent energy efficiency program efforts have focused on understanding the behavior and decision-making of individuals and organizations with respect to the design, adoption, and use of energy efficiency actions and on using that knowledge to help accelerate the implementation of energy efficiency activities. The proceedings of the 2007 Behavior, Energy and Climate Change Conference provide information on these approaches. See <http://ciee.ucop.edu/>.

For education and training programs, evaluations focus on documenting the degree to which the programs are achieving their desired effects within the markets targeted by the program, which is educating and training people on energy efficiency. The primary mechanisms for this type of evaluation are surveys and focus groups. The following are examples of information topics that may be collected as part of surveys and focus groups (paraphrased from the *California Protocols*):

- Information and education program evaluation topics:
  - Number and percent of customers reached or made aware.
  - Number and percent of customers reached who take recommended actions.
  - Number and type of actions taken as a result of the program.
  - Changes in awareness or knowledge by topic or subject area, by type of customer targeted.
  - Customer perception of the value of the information and/or education received.
  - Elapsed time between information exposure and action(s) taken by type of customer targeted.

  - Attribution of cause for actions taken when multiple causes may be associated with the actions taken.
  - Influence of the program on dealers, contractors, and trade allies.
  - Effects of the program on manufacturers and distributors.
- Training program evaluation topics:
  - Pre-program level of knowledge to compare with post-program levels.
  - The specific knowledge gained through the program.
  - The relevance and usefulness of the training as it relates to the participants' to specific needs and opportunities to use the information.
  - Future opportunities and plans for incorporating the knowledge gained into actions or behaviors that provide energy impacts.
  - Whether participants would recommend the training to a friend or colleague.
  - Participant recommendations for improving the program.

Note that programs with large training efforts, or programs designed solely for training, should have evaluation designs that are mindful of the rich literature and methods on evaluating training programs that are available from the larger evaluation community.

Appendix D: Uncertainty[1]

This appendix provides an introduction on how uncertainty is defined, as well as an overview of the range of factors that contribute to uncertainty and the impact of each of these. This discussion's target audience is evaluators who need an introduction to uncertainty and managers responsible for overseeing evaluations, such as government, regulatory agency staff, and utility staff responsible for energy efficiency evaluations. The reader will gain a solid foundation for understanding key concepts and determining evaluation strategies for identifying and mitigating uncertainty, as well as the ability to review, as needed, more technical and detailed discussions of each source of uncertainty and its mitigation.[2]

D.1 Sources of Uncertainty

Uncertainty is a measure of the "goodness" of an estimate. Without some measurement of uncertainty, it is impossible to judge an estimate's value as a basis for decision-making: uncertainty is the amount or range of doubt surrounding a measured or calculated value. Any report of gross or net program savings, for instance, has a halo of uncertainty surrounding the reported value relative to the true gross or net savings (which are not known). Defined this way, uncertainty is an overall indicator of how well a calculated or measured value represents a true value.

Program evaluation seeks to reliably determine energy and demand *savings* (and, potentially, non-energy benefits) with some reasonable accuracy. This objective can be affected by:

- **Systematic sources of error,** such as measurement error, non-coverage error, and non-response error.
- **Random error**—error occurring by chance, attributable to using a population sample rather than a census to develop the calculated or measured value.

The distinction between systematic and random error is important because different procedures are required to identify and mitigate each. The amount of random error can be estimated using statistical tools, but other means are needed for systematic error. While additional investment in the estimation process reduce both types of error, tradeoffs between evaluation costs and reductions in uncertainty are inevitably required.

D.1.1. Sources of Systematic Error

Systematic errors potentially occur from the way data are:

- **Measured.** At times, equipment used to measure consumption may not be completely accurate. Human errors (e.g., errors in recording data) can also cause this type of error. Measurement error is reduced by investing in more accurate measurement technology and more accurately recording and checking data. The magnitude of such errors is often not large enough to warrant concern in a program evaluation and is largely provided by manufacturer's specifications. In most applications, this error source is ignored, particularly when data sources are utility-grade electricity or natural gas meters. However, other types of measurements, such as flow rates in water or air distribution systems, can have significant errors.
- **Collected.** If some parts of a population are not included in the sample, non-coverage errors result, and the value calculated from the sample might not accurately represent the entire population of interest. Non-coverage error is reduced by investing in a sampling plan that addresses known coverage issues. For instance, a survey implemented through several modes, such as phone and Internet, can sometimes address known coverage issues. Non-response errors

occur when some portion or portions of the population, with different attitudes or behaviors, are less likely to provide data than are other portions. For a load research or metering study, if certain types of households are more likely to refuse to participate or if researchers are less likely to be able to obtain required data from them, the values calculated from the sample will understate the contribution of this portion of the population and over-represent the contribution of sample portions more likely to respond. In situations where the under-represented portion of the population has different consumption patterns, non-response error is introduced into the value calculated from the sample. Non-response error is addressed through investments that increase the response rate, such as incentives and multiple contact attempts.

- **Described (modeled).** Estimates are created through statistical models. Some are fairly simple and straightforward (e.g., estimating the mean), and others are fairly complicated (e.g., estimating response to temperature through regression models). Regardless, errors can occur due to the use of the wrong model, assuming inappropriate functional forms, inclusion of irrelevant information, or exclusion of relevant information. For example, in determining energy savings, a researcher may be required to adjust measured energy use data to make comparisons with a baseline. This process can introduce systematic errors.

D.1.2 Sources of Random Error

Whenever a *sample* of a population is selected to represent the population itself—whether the sample is of appliances, meters, accounts, individuals, households, premises, or organizations—there will be some amount of *random sampling error*. The sample selected is only one of a large number of possible samples of the same size and design that could have been selected from that population. For each sample, values calculated will differ from the other potential samples simply because of the element of chance in choosing particular elements. This variability is termed random sampling error. Random sampling error, unlike the systematic errors discussed above, can be estimated using statistical tools (assuming the sample was drawn randomly).

When the time savings actually take place is also essential—another layer of sampling error. Typically, what (or who) is sampled and when they are sampled (e.g., metering energy consumption over one week, metering 5 percent of impacted equipment) introduces uncertainty.

Altering sample design can reduce uncertainty from random sampling error (for instance, increasing the number of elements sampled or changing the way elements are grouped together prior to sampling). As expected, random error and sampling costs are inversely proportional in most instances.

In addition to random sampling error, random measurement error may be introduced by other factors, such as respondents' incorrectly recalling dates or expenses, or other differences in a respondent's mood or circumstances that affect how they answer a question. These other types of random measurement error are generally assumed to "even out," so that they do not affect the mean or point estimate, but only increase the variability. For this reason, researchers generally do not attempt to quantify the potential for random measurement error in the data.

D.2 Energy Efficiency Evaluation Uncertainty

The biggest challenge in evaluating energy efficiency programs is a lack of direct measurement. Energy savings are what *did not happen*, but energy consumption is actually what is measured. The difference between energy consumption and what energy consumption *would have been* had energy efficiency measures not been installed provides a measure of energy savings. Savings computation therefore involves comparing measured energy data and a calculation of "adjustments" to convert both measurements to the same set of operating conditions (i.e., a baseline). Both measurement and adjustment processes introduce uncertainty.

These processes produce statistical "estimates" with reported or expected values and some level of variability.

In other words, true values cannot be known; only estimates can be made, with some level of uncertainty. Physical measurements and statistical analyses are based on estimation of central tendencies (mean, median, mode) and associated quantification of variations (standard deviation, standard error, variance).

Because uncertainty arises from many different sources, it is usually difficult to identify and quantify the effect of all potential sources. Research reports often identify only uncertainty arising from random sampling error, because this source of error is usually the easiest component to quantify. Convenient measures, such as confidence intervals and statistical significance tests, are available to provide quantitative estimates of the uncertainty. Uncertainty attributable to forms of systematic error does not have a single comparable measure to provide a parsimonious estimate of uncertainty. Rather, these sources are specific to individual studies, depending on equipment used, research staff, or research and data collection procedures employed. To assess uncertainty from systematic sources, evaluators must address the rigor of evaluation procedures.

Evaluating uncertainty is an ongoing process that can consume time and resources. It may also require the services of specialists familiar with data analysis techniques, further data collection, or additional equipment. Reducing errors usually increases evaluation costs. Thus, improved accuracy should be justified by the value of the improved information.

D.3 Statistical Terms

While studying a phenomenon at the population level (a census) produces greater accuracy, the cost is almost always prohibitive. If properly designed, samples can provide accurate estimates at a greatly reduced cost. Statistics are mathematical methods that, applied to sample data, can help make inferences about whole populations and aid decisions in the face of uncertainty.

For any value calculated from a sample, a set of descriptive statistics, such as the mean, standard deviation, standard error, and a confidence interval, can be calculated. Standard deviation is a measure of variability showing the extent of dispersion around the mean. In normally distributed data, about 68 percent of observations are within one standard deviation of the mean; so a large standard deviation indicates greater dispersion of an individual observation from each sample member, while a smaller standard deviation indicates less dispersion. Based on the amount of variability and standard deviation, a confidence interval can be calculated.

To communicate evaluation results credibly, outcomes need to be expressed with their associated variability. *Confidence* refers to the probability the estimated outcome will fall within some level of *precision*. Statement of precision without a statement of *confidence* proves misleading, as evaluation may yield extremely high precision with low confidence or vice versa. For example, after metering a sample of impacted equipment, one may estimate average savings as 1,000 kWh. This is an *estimate* of the *true average* savings. Further, one may able to state the true average is within ±1 percent of the estimate (precision), but only be 30 percent confident that is the case. Alternatively, one may be 99 percent confident the true average savings are within ±50 percent of the estimate of 1,000 kWh.

If the estimated outcomes are large relative to the variation, they tend to be statistically significant. On the other hand, if the amount of variability is large relative to the estimated outcome, one is unable to discern if observed values are real or simply random. In other words, when variability is large, it can lead to precision levels that are too large (e.g., more than ±100 percent) for observed estimates (e.g., estimated savings) to be meaningful. In an extreme example, if the observed average is 1,000 kWh and the associated precision is ±150 percent, true average savings are somewhere between negative 500 kWh (which means the measure actually caused consumption to increase) and 1,500 kWh.

To formalize these relationships, evaluators use the *t* statistic test. The *t* statistic is a measure of a statistical estimate's reliability. When the parameter estimate, such as mean kW savings, is small relative to its associated variability, the *t* statistic value is low. For the rest of this section example values are presented using a 95 percent level of confidence, for which the critical value of *t* is approximately 2. With a required 95 percent confidence level if the *t* statistic is less than 2, the estimated parameter is considered not reliable.

Confidence intervals are a convenient way of expressing the potential random sampling error for an estimate. Confidence intervals are calculated by multiplying the estimated standard error by a value based on the *t* statistic and adding or subtracting this number from the estimate. For example, once average savings are estimated, true average savings are bracketed in the following confidence interval:

$$\text{estimated average savings} - t(SE_{\text{savings}}) \le \text{true average savings} \le \text{estimated average savings} + t(SE_{\text{savings}})$$

Table D-1 summarizes the statistical terms useful for in assessing uncertainty. (The table provides an easy reference, not a guide for computations.)

For example, assume that 12 monthly energy bills total 48,000 kWh. Estimated average annual consumption is:

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{48{,}000}{12} = 4{,}000$$

The *variance* is:

$$S^2 = \frac{\sum (Y_i - \bar{Y})^2}{n-1} = 4{,}488{,}417 \text{ kWh}^2$$

The *standard deviation* is:

$$s = \sqrt{S^2} = \sqrt{4{,}488{,}417} = 2{,}118\ kHw$$

The *standard error* is:

$$SE = \frac{s}{\sqrt{n}} = \frac{2{,}118}{\sqrt{12}} = 611\ kHw$$

Thus, at a 95 percent confidence level, the *absolute precision* is approximately:

$$t \times SE = 2 \text{ X } 611 = 1{,}222\ kWh$$

At a 95 percent confidence level, the *relative precision* is:

$$\frac{t \times SE}{estimate} = \frac{1{,}222}{4{,}000} = 30\%$$

Table D-1. Summary of Statistical Terms

Mean ($\bar{Y}$)	The mean is determined by adding up individual data points and dividing by the total number of these data points.	$\bar{Y} = \frac{\sum Y_i}{n}$
Variance (S^2)	The extent to which observed values differ from each other. Variance is found by averaging the squares of individual deviations from the mean. Deviations from the mean are squared simply to eliminate negative values.	$S^2 = \frac{\sum Y_i\ (Y_i - \bar{Y})^2}{n-1}$
Standard Deviation (*s*)	This is simply the square root of the variance. It brings the variability measure back to the units of the data (e.g., while variance units are in kWh^2, the standard deviation units are kWh).	$s = \sqrt{S^2}$
Standard Error (*SE*)	The standard deviation divided by the square root of the total number of observations. *SE* is the measure of variability used in assessing precision and confidence for the true value of the estimate.	$SE = \frac{s}{\sqrt{n}}$
Coefficient of Variance (*cv*)	Defined as the standard deviation of the readings divided by the mean, this is used in estimating sample sizes.	$CV = \frac{s}{\bar{Y}}$
Absolute Precision	Computed from standard error using a *t* value.	$t * SE$
Relative Precision	The absolute precision divided by the estimate.	$\frac{t * SE}{Estimate}$

That is, based on observing a sample of 12 months, average monthly consumption is estimated at 4,000 kWh. There is a 95 percent confidence that the *true* mean monthly consumption lies between 2,778 and 5,222 kWh. It can be said with 95 percent confidence that the *true* mean value is 4,000 ±30 percent.[3]

D.4 Mitigating Random/Sampling Error

In most evaluations, the entire population of interest (e.g., all small commercial customers participating in a program) cannot be accessed, either because the population is too large or the measurement process is too expensive or time-consuming to allow more than a small segment of the population to be observed. Therefore, decisions about a population are based on a small amount of sample data.

For example, suppose an evaluator is interested in the proportion of program participants installing a particular measure. The fairly large program has a population of about 15,000 participants. The parameter of interest is the proportion of participants actually installing the measure (called π).

The evaluator conducts a survey using a random sample of participants. Each participant is asked whether or not they installed the measure. The number (call it n) of participants surveyed will be quite small relative to the population's size. Once these participants have been surveyed, the proportion installing the measure will be computed. This proportion is called a statistic (in this case, it is called p). The evaluator can reasonably assume p will not equal π (an exact match would be extremely unlikely). At a minimum, p involves random sampling error or "the luck of the draw." The difference between the observed p and unobserved π is the sampling error. As long as sampling is used, there will be sampling error.

The most direct way to reduce sampling error is to increase the sample's size. Most research consumers are familiar with this underlying principle. For any given population and confidence level, the larger the sample, the more precise estimates will be.

Evaluation research adopts conventions about sample sizes for particular types of projects. Prior research (or in some cases, requirements set by a regulatory authority) should be the first place to turn for appropriate sample sizes. The next question is whether relationships in prior studies seem likely to exist but have not been borne out by research. This might point toward the need to invest in a larger-than-conventional sample size.

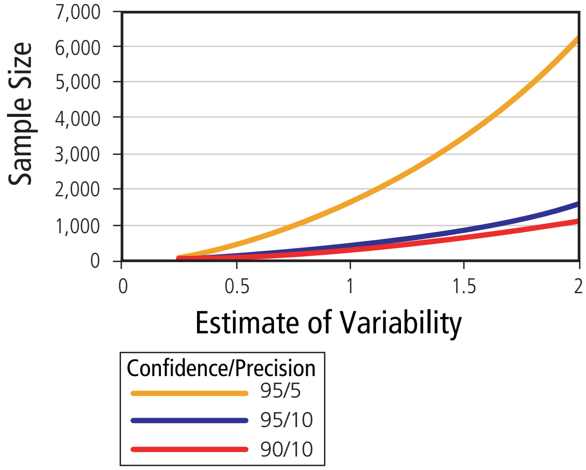
The other way to reduce sampling error is to improve the sampling design. In general, the design with the smallest random error is a simple random sample in which each population element has an equal probability of being selected. There are important reasons why a deviation from this design represents an overall improvement in results. For example, using a stratified sampling design (rather than treating all lighting areas as if they were all part of the same "population") divides populations into homogenous strata prior to sampling, greatly reducing overall sampling error. Researchers should justify stratification or clustering of the sample and address the impact on sampling error.

As noted, sampling error can be minimized by increasing the fraction of the population sampled, obviously at an increased cost. Several issues are critical in optimizing sample sizes. The following steps should be followed in setting the sample size.

1. **Select a homogeneous population.** For sampling to be cost-effective, measured "units" should be expected to be the same for the entire population. If there are two different types of units in the population, they should be grouped and sampled separately. For example, when designing a sampling program to measure the operating periods of room lighting controlled by occupancy sensors, rooms occupied more or less continuously (e.g., multiple-person offices) should be sampled separately from those that are only occasionally occupied (e.g., meeting rooms). The size of the sample needed to achieve a certain level of precision and confidence is sensitive to the amount of variability. Figure D-1 presents a hypothetical case. The horizontal axis shows an estimate of variability (in this case, the *cv*). The vertical axis shows the sample size needed to achieve different levels of

confidence and precision. Each of the lines shows the relationship between variability and sample size. The higher the confidence and precision requirement, the steeper the line, indicating higher sensitivity to the measure of variability. This clearly displays the need for homogeneous groups. A homogeneous group is defined as a group with low variability in whatever is being measured (e.g., hours of use).

Figure D-1. Sample Size Selection for Different Levels of Confidence and Precision



Provided by Khawaja and Baumgarten of Quantec, LLC.

In addition to placing the population in homogenous groups, the evaluator needs to set the acceptable levels of precision and confidence. A conventional approach (for example, as defined for certain California energy efficiency evaluations) is to design sampling to achieve a 90 percent *confidence* level and ±10 percent *precision*. Figure D-1 illustrates the impact of selecting confidence and precision levels. For example, in the hypothetical situation illustrated, the sample size needed for a *cv* of 1.0 varies from 270 for 90/10 to over 1,500 for 95/5. This may translate to a difference of thousands of dollars in sampling costs: improving *precision* from ±20 to ±10 percent would require a fourfold increase in sample size, while improving it to ±2 percent would require a hundredfold increase in sample size. This is due to a sample error inversely proportional to $\sqrt{n}$. Thus, selecting the appropriate sampling criteria requires balancing accuracy requirements and the risk of higher uncertainty with costs associated with less uncertainty.

2. **Decide on the level of disaggregation.** It is necessary to establish whether the *confidence* and *precision* level criteria should be applied to the measurement of all components or to various subgroups of components. If a project includes several measures installed in different building types, the evaluator must decide whether the confidence and precision apply at the project level, measures level, end-use level, and so on. Going from measure-level criteria to overall project-level criteria requires larger samples. However, one large sample covering an entire project may still be smaller than several smaller samples at the measure level. As there are no hard and fast rules, different sampling designs need to be examined, and those optimally balancing the precision and cost should be selected. Whatever that final selection, it should be clearly defined in an evaluation plan along with the rationale behind the selection.

3. **Calculate initial sample size.** Based on the information above, an initial estimate of the overall sample size required to meet the research goals can be determined using the following equation:

$$n_o = \frac{z^2 * cv^2}{e^2}$$

where:

n_o	is the initial estimate of the required sample size before sampling begins.
cv	is the *coefficient of variance*, defined as the *standard deviation* of the readings divided by the *mean*. Until the actual *mean* and *standard deviation* of the population can be estimated from actual samples, 0.5 is often accepted as an initial estimate for *cv*. The more homogenous the population, the smaller the *cv*.
e	is the desired level of *precision*.
z	is the standard normal distribution value for the desired *confidence* level. For example, *z* is 1.96 for a 95 percent *confidence* level (1.64 for 90 percent, 1.28 for 80 percent, and 0.67 for 50 percent *confidence*).

For 90 percent *confidence* with 10 percent *precision* and a *cv* of 0.5, the initial estimate of required sample size (n_o) is

$$n_o = \frac{1.64^2 \times 0.5^2}{0.1^2} = 67$$

Values from previous *cv* studies may be used if available. It may also be desirable to conduct a study with a small sample for the sole purpose of estimating *cv*.

4. **Adjust initial sample size estimate for small populations.** The necessary sample size can be reduced if the entire population being sampled is no more than 20 times the size of the sample. For the initial sample size example above (number = 67), if the population *(N)* from which it is sampled is only 200, the population is only 3 times the size of the sample. Therefore the "finite population adjustment" can be applied. This adjustment reduces the sample size *(n)* as follows:

$$n = \frac{n_o N}{n_o + N}$$

Applying this finite population adjustment to the above example reduces the sample size *(n)* required to meet the 90 percent/±10 percent criterion to 50.

D.5 Mitigating Systematic Error

Many evaluation studies do not report any uncertainty measures besides a sampling error–based confidence interval for estimated energy or demand savings values. This is misleading because it suggests the confidence interval describes the total of all uncertainty sources (which is incorrect) or that these other sources of uncertainty are not important relative to sampling error. Sometimes uncertainty due to measurement and other systematic sources of error can be significant.

Measurement error can result from inaccurate mechanical devices, such as meters or recorders, as well as from inaccurate recording of observations by researchers or inaccurate responses to questions by study participants. Of course, basic human error occurs in taking physical measurements or conducting analyses, surveys, or documentation activities. For mechanical devices such as meters or recorders, it is theoretically possible to perform tests with multiple meters or recorders of the same make and model to indicate the variability in measuring the same value. However, for meters and most devices regularly used in energy efficiency evaluations, it is more practical to either use manufacturer and industry study information on the likely amount of error for any single piece of equipment or use calibration data.

Assessing the level of measurement error for data obtained from researchers' observations or respondents' reports is usually a subjective exercise, based on a qualitative analysis. The design of recording forms or questionnaires, the training and assessment of observers and interviewers, and the process of collecting data from study participants are all difficult to quantify. It is possible, however, to conduct special studies of a participant subsample to validate each of these processes. For example, it is possible to have more than one researcher rate the same set of objects, or to conduct short-term metering of specific appliances for a subsample to verify information about appliance use. Participants can also be reinterviewed to test the answer to the same question at two different times, and pretests or debriefing interviews can be conducted with participants to determine how they interpreted specific questions and constructed their responses. Such special studies can be used to provide an assessment of the uncertainty potential in evaluation study results.

Another challenge lies in estimating the effect of excluding a portion of the population from a sample (sample non-coverage) or of the failure to obtain data from a certain portion of the sample (non-response). Data needed to assess these error sources are typically the same as those needed to resolve errors in the first place—but these data are usually unavailable. However, for both non-coverage and non-response, it is possible to design special studies to estimate the uncertainty level introduced. For studies whose sample design did not include a particular portion of the population (such as a geographical area or respondents living in a certain type of housing), it is possible to conduct a small-scale study on a sample of the excluded group to determine the magnitude and direction of differences in calculated values for this portion of the population. In some situations, such as a survey, it is also possible to conduct a follow-up study of a sample of members for whom data were not obtained. This follow-up would also provide data to determine if non-respondents were different from respondents, as well as an estimate of the magnitude and direction of the difference.

Determining steps needed to mitigate systematic error is a more complex problem than mitigating random error as various sources of systematic error are often specific to individual studies and procedures. To mitigate systematic error, evaluators typically need to invest in additional procedures (such as meter calibration, a pretest of measurement or survey protocols, a validation study, or a follow-up study) to collect additional data to assess differences between participants who provided data and those who did not.

To determine how rigorously and effectively an evaluator has attempted to mitigate sources of systematic error, the following should be examined:

1. Were measurement procedures, such as the use of observational forms or surveys, pretested to determine if sources of measurement error could be corrected before the full-scale study was fielded?
2. Were validation measures, such as repeated measurements, inter-rater reliability, or additional sub-sample metering, used to validate measurements?
3. Was the sample frame carefully evaluated to determine what portions of the population, if any, were excluded in the sample and, if so, what steps were taken to estimate the impact of excluding this portion of the population from the final results?
4. Were steps taken to minimize the effect of non-response in surveys or other data collection efforts? If non-response appears to be an issue, were steps taken to evaluate the magnitude and direction of potential non-response bias?
5. Has the selection of formulas, models, and adjustments been conceptually justified? Has the evaluator tested the sensitivity of estimates to key assumptions required by the models?
6. Did trained and experienced professionals conduct the work, and was it checked and verified by a professional other than the one conducting the initial work?

D.6 Addressing More Complex Uncertainty

This discussion has assumed that uncertainty arises from variation in one variable (e.g., hours of use or level of consumption). Often, uncertainty is caused by variability in several components in a savings estimation equation. For example, total savings may be the sum of savings from different components:

$$savings = savings_1 + savings_2 + \ldots + savings_p$$

Where total savings are the result of lighting, cooling, and so on. Each savings component is likely to have some variability of its own. Combining savings into the total requires the evaluator to also combine the variability associated with the different estimates. Components must be *independent* to use the suggested methods for combining uncertainties. Independence means whatever random errors affect one component are unrelated to the affecting other components. The standard error of reported savings can be estimated by:

$$SE(savings) = \sqrt{SE(savings_1)^2 + SE(savings_2)^2 + \ldots\ldots + SE(savings_p)^2}$$

Savings can also be estimated as the difference between baseline and post-installation energy use. The *standard error* of the difference (*savings*) is computed as:

$$SE(savings) = \sqrt{SE(adjusted\ baseline)^2 + SE(reporting\ period\ energy)^2}$$

At times, the savings estimate is a *product* of several independently determined components (i.e., *savings*=$C_1 \times C_2 \times \ldots \times C_p$); in that case, the *relative* standard error of the *savings* is given approximately by:

$$\frac{SE(savings)}{savings} \approx \sqrt{\left(\frac{SE(C_1)}{C_1}\right)^2 + \left(\frac{SE(C_2)}{C_2}\right)^2 + \ldots\ldots + \left(\frac{SE(C_p)}{C_p}\right)^2}$$

A good example of this is the determination of lighting savings as:

$$savings = \Delta \text{ Watts} \times \text{Hours}$$

The relative standard error of savings will be computed using the above formula as follows:

$$\frac{SE(savings)}{savings} = \sqrt{\left(\frac{SE(\Delta\ watts)}{\Delta\ watts}\right)^2 + \left(\frac{SE(hours)}{hours}\right)^2}$$

If savings at a particular hour are what is needed, the affected end-use must be metered hourly. The estimated average is energy use in that particular hour. The variability measure is the usage observed at that hour, and the sampling unit is the number of hours to be metered. Metering periods must account for weather and other seasonal variations, and metered hours must include a sample of different use patterns. In other words, sampling becomes more complex as an evaluator needs to estimate a sample size in number of hours per usage pattern as well as the number of impacted end-uses to be metered.

In many cases, the estimate of uncertainty attributable to systematic errors will have to be stated in qualitative terms. However, it is important to recognize these error sources may be significant. As a result, relying only on confidence intervals and standard errors to express uncertainty may be very misleading.

D.7 Monte Carlo Methods

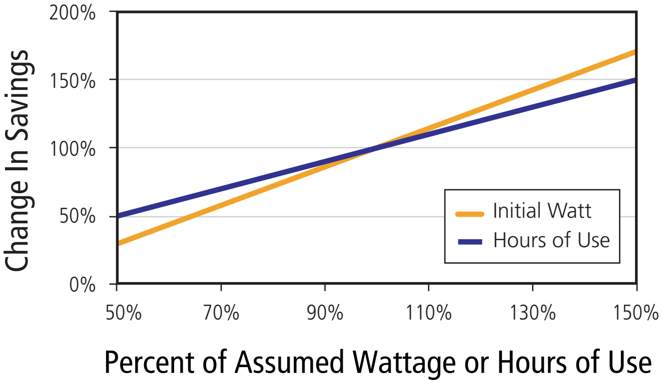
Earlier sections discuss uncertainty as a range of values surrounding a point value that has been arrived at directly through a measurement process. Monte Carlo methods arrive at uncertainty in a different way: by simulating reality using chance (hence the gambling reference) and a model of factors contributing uncertainty to the outcome we are examining.

A "Monte Carlo" method can be any technique using random numbers and probability to develop a simulation to solve a problem. Monte Carlo methods are used in many fields, such as chemistry and physics, as well as in studying how to best regulate the flow of traffic on highways or in risk management for businesses and organizations. They are used when a large number of uncertainty sources exist in the inputs, and direct measurements of outcomes are not possible. This appendix describes Monte Carlo simulation used to estimate a population distribution.

Monte Carlo techniques are perfectly viable alternatives when problems are too complex (i.e., too many factors are involved in computing savings). Assessing the importance of individual components is often the best first step in assessing uncertainty.

For example, the Monte Carlo method could be applied for a simple lighting retrofit. Simply stated, such analysis begins by allowing these factors (e.g., hours of use) to vary from plausible lows to plausible highs. The impact on final savings values are observed, then the impact of initial wattage is investigated by allowing its value to vary between plausible lows and highs. The factor that has a higher impact on final savings may be the one worthy of further research. Figure D-2 shows a hypothetical case in which the initial wattage level and hours of use are allowed to vary independently from 50 percent of assumed or most likely values to 150 percent. Savings are estimated for all these variations. The vertical axis shows the change in savings as percentage terms. In this example, the initial wattage has a steeper curve, indicating higher sensitivity of final savings estimates.

Figure D-2. Hypothetical Analysis—Lighting Project



Provided by Khawaja and Baumgarten of Quantec, LLC.

Some commercially available software (e.g., Crystal Ball™) uses a Monte Carlo simulation to perform this type of analysis. These models are usually built using spreadsheets and are organized so ranges can be entered by the evaluator for each input variable needed to perform the sensitivity analysis shown above. Monte Carlo simulation is a very flexible technique, widely used for assessment of risk analysis in various fields.

In the example, hours of use and initial wattage levels were allowed to vary independently one at a time. Using Monte Carlo tools, the evaluator can allow factors to vary concurrently. Thus, Monte Carlo simulation can be as simple or complex as the user requires.

An extended example provides the best explanation of how this approach might be used in evaluating an energy efficiency program. Suppose energy savings from a residential air conditioning upgrade program are assumed to be a function of the net number of participants (subtracting free riders), multiplied by the average savings from each installed air conditioner. The savings from each installed air conditioner is a function of the average difference in SEER, relative to the old unit and a behavioral component related to thermostat settings also incorporated in the program.

Given these assumptions:

- There is uncertainty about the number of free riders. Estimates yield a 25 percent probability that 10 percent of participants are free riders, a 50 percent probability that 20 percent of participants are free riders, and a 25 percent probability that 30 percent of participants are free riders.
- There is also uncertainty regarding the average difference in SEER. That difference has not been directly measured through surveys, but the distribution of SEER values for qualifying air conditioners is known as is the average SEER of currently installed air conditioners. It is not known whether program participants are different than the general population. Estimates show a 25 percent probability that the average difference in SEER values is SEER 1.5, a 50 percent probability that the average difference is SEER 2.0, and a 25 percent probability that the average difference is SEER 2.5.
- Finally, uncertainty exists regarding the behavioral component. It is believed that there is a 50 percent probability that the average effect of the campaign has no change in thermostat settings, a 40 percent probability that the average effect reduces settings by 0.5 degrees, and a 10 percent probability that the average effect reduces settings by 1.0 degree.

This example models 27 possible scenarios (all possible combinations of the three factors, i.e., 33) and can calculate a savings for each state. Using the probability of each state, it is possible to estimate a probability distribution for program savings, including a mean, standard deviation, and confidence intervals. For instance, the probability that actual savings are at the peak estimate (where free ridership is low, SEER difference is high, and thermostat setting is reduced by 1.0 degree) is 0.25 × 0.25 × 0.10 = 0.00625, or 0.625 percent.

So far this example does not involve chance because the example has only 27 possible scenarios. As the number of factors or states increases, it becomes impossible to calculate savings for every possible combination. If there are 10 uncertainty factors, with each having 10 possible states, there are 10^{10} or 10 billion possible combinations. To estimate uncertainty, the population of scenarios is simulated using random number generators and multiple samples of a reasonable size are drawn—for instance, 1,000 samples of 1,000 scenarios. For each sample, mean program savings are calculated. Given the laws of probability, the average of the 1,000 samples (i.e., the average of the averages) is a good point estimate of energy savings, and the distribution around that estimate is normally distributed and provides a good estimate of uncertainty surrounding the estimate.

The key caution about Monte Carlo analysis is that, as with all simulations, poor assumptions built into the model can yield inaccurate estimates of the true uncertainty surrounding an estimate. Nevertheless, in fields where interrelations are very complex or direct measurements impossible, Monte Carlo analysis can yield useful uncertainty estimates.

D.8 Notes

1. This appendix was prepared by Dr. M. Sami Khawaja, President, Quantec, LLC, and Dr. Bob Baumgartner, Principal, PA Consulting.

2. This appendix assumes readers are *not* trained statisticians and does not aim to provide the reader with all of the tools, formulas, and programs to calculate measures of uncertainty.

3. These are approximations; in actual applications the exact value of *t* would be used. Also, for the same example, if the desired confidence were 90 percent, the relative precision estimate would be approximately ±25 percent. At 80 percent, it would be about ±20 percent.

Appendix E: Resources

This appendix provides an overview of the key documents used in the development of this Guide. These documents were developed with information gathered over the last several decades of energy efficiency program implementation and evaluation. They can be considered the current primary resources for efficiency program evaluation and project M&V, and thus the basis for the definitions, approaches, and issues adopted and explained in this Guide.

E.1 Background

The information in this Guide is a summary of definitions, approaches, and issues that have developed over the last 30 years of energy efficiency program implementation and evaluation. This experience and expertise is documented in numerous guides, protocols, papers, and reports. From a historical perspective, many of the basic references on energy and energy efficiency impact evaluations were written in the 1980s and 1990s. There are two reference documents in the public domain that provide a historical perspective and solid fundamentals:

- Violette, D.M. (1995). *Evaluation, Verification, and Performance Measurement of Energy Efficiency Programmes*. Prepared for International Energy Agency.
- Hirst, E., and J. Reed, eds. (1991). *Handbook of Evaluation of Utility DSM Programs.* Prepared for Oak Ridge National Laboratory ORNL/CON-336.

However, most of the early reference documents are not easily available to the general public (i.e., they are not posted on the Web).

E.2 Primary Impact Evaluation Resources

The key documents used in the development of this Guide are available via the Web and are presented in this section; they can be considered the current primary resources for efficiency program evaluation and project M&V. These documents are well-established *project* M&V guides and *program* evaluation protocols. They constitute the core M&V guidance documents used for energy efficiency projects in the United States and many other countries.

- **2007 International Performance Measurement and Verification Protocol (IPMVP).** The IPMVP provides an overview of current best practices for verifying results of energy efficiency, water, and renewable energy projects in commercial and industrial facilities. Internationally, it is the most recognized M&V protocol for demand-side energy activities. The IPMVP was developed with DOE sponsorship and is currently managed by a nonprofit organization[1] that continually maintains and updates it.

  The IPMVP provides a framework and definitions that can help practitioners develop M&V plans for their projects. It includes guidance on best practices for determining savings from efficiency projects. It is not a "cookbook" of how to perform specific project evaluations; rather, it provides guidance and key concepts that are used in the United States and internationally. The IPMVP is probably best known for defining four M&V Options for energy efficiency projects. These Options (A, B, C and D) differentiate the most common approaches for M&V and are presented in Chapter 5.

  Reference: Efficiency Valuation Organization (2007). *International Performance Measurement and Verification Protocol*. <http://www.evo-world.org>

- **2000 FEMP M&V Guidelines.[2]** The purpose of this document is to provide guidelines and methods for measuring and verifying the savings associated with federal agency performance contracts. It contains procedures and guidelines for quantifying the savings resulting from energy efficiency equipment, water conservation, improved operation and maintenance, renewable energy, and cogeneration projects.

  References: U.S. Department of Energy (2000). *M&V Guidelines: Measurement and Verification for Federal Energy Projects*. Version 2.2. <http://ateam.lbl.gov/mv/docs/26265.pdf>

  U.S. Department of Energy (2002). *Detailed Guidelines for FEMP M&V Option A*. <http://ateam.lbl.gov/mv/docs/OptionADetailedGuidelines.pdf>

- **2002 ASHRAE Guideline 14 Measurement of Energy and Demand Savings.[3]** ASHRAE is the professional engineering society that has been the most involved in writing guidelines and standards associated with energy efficiency. Compared to the FEMP M&V Guidelines and the IPMVP, Guideline 14 is a more detailed technical document that addresses the analyses, statistics, and physical measurement of energy use for determining energy savings.

  Reference: American Society of Heating, Refrigerating, and Air-Conditioning Engineers (2002). *Guideline 14 on Measurement of Demand and Energy Savings*.

In addition, in terms of energy efficiency *program* protocols, two documents are often cited as standards in the United States for energy efficiency evaluation:

- California Public Utilities Commission (2006). *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals*. <http://www.calmac.org/publications/EvaluatorsProtocols_Final_AdoptedviaRuling_06-19-2006.pdf>
- California Public Utilities Commission (2004). *The California Evaluation Framework*. <http://www.calmac.org/publications/California_Evaluation_Framework_June_2004.pdf>

These documents provide a great deal of information on evaluation options and principles for impact, process, and market evaluations of a wide variety of energy efficiency program types. In many respects, they are a more detailed version of this Guide. Along with many other evaluation reports and guidance documents, they can be found at two Web-accessible databases:

- CALifornia Measurement Advisory Council (CALMAC): <http://www.calmac.org>.
- The Consortium for Energy Efficiency's Market Assessment and Program Evaluation (MAPE) Clearinghouse: <http://www.cee1.org/eval/clearinghouse.php3>.

Readers can also look at the Proceedings of the IEPEC Conference (<http://www.iepec.org>) and ACEEE Summer Studies (<http://www.aceee.org>), where there are shorter (10- to 12-page) examples of evaluations (versus the 100+ pages for a typical evaluation study).

Three other important program guides are:

- International Energy Agency (2006). *Evaluating Energy Efficiency Policy Measures & DSM Programmes*. <http://dsm.iea.org>
- U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy (2003). EERE Program Analysis and Evaluation. In *Program Management Guide: A Reference Manual for Program Management*. <http://www1.eere.energy.gov/ba/pdfs/pm_guide_chapter_7.pdf>
- U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy (2007). *Impact Evaluation Framework for Technology Deployment Programs*. Prepared by J. Reed, G. Jordan, and E. Vine. <http://www.eere.energy.gov/ba/pba/km_portal/docs/pdf/2007/impact_framework_tech_deploy_2007_main.pdf>

Another important resource is the Database for Energy Efficient Resources (DEER). Sponsored by the California Energy Commission and CPUC, DEER provides estimates of energy and peak demand savings values, measure costs, and effective useful life. CPUC has designated

DEER its source for deemed and impact costs for program planning. The current version (October 2005) has more than 130,000 unique records representing over 360 unique measures within the DEER dataset. The data are presented as a Web-based searchable data set: <http://www.energy.ca.gov/deer/index.html>.

For calculating avoided emissions, several publications prepared as part of the Greenhouse Gas Protocol Initiative were consulted. The Initiative is a multi-stakeholder partnership of businesses, non-government organizations (NGOs), governments, and others convened by the WRI and the WBCSD. The Initiative's mission is to develop internationally accepted accounting and reporting protocols for corporate emissions inventories and greenhouse gas mitigation projects and to promote their use by businesses, policy makers, NGOs, and other organizations. It consists of three GHG accounting modules, as well as outreach activities. The accounting modules are:

- **Corporate Accounting and Reporting Standard.** Standards, guidance, and Web-based calculation tools to help companies, regulators and others develop an organization-wide greenhouse gas emissions inventory.
- **GHG Project Accounting and Reporting Protocol.** Requirements and guidance for quantifying reductions from greenhouse gas mitigation projects, such as those used to offset emissions or to generate credits in trading programs.
- **Guidelines for Quantifying GHG Reductions from Grid-Connected Electricity Projects.**

These documents are available at <http://www.wri.org/climate/>.

Another series of greenhouse gas guides is the International Organization for Standardization (ISO) 14064 series. There are three parts to the ISO 14064 standards:

- **ISO 14064-1,** which specifies principles and requirements at the organization level for the design, development, management, maintenance, and verification of an organization's GHG inventory.
- **ISO 14064-2,** which specifies principles and requirements and provides guidance at the project level for quantifying and reporting activities intended to cause GHG emission reductions or removal enhancements.
- **ISO 14064-3,** which specifies principles and requirements and provides guidance for those conducting or managing the validation and/or verification of GHG assertions, such as the validation or verification of an organization's GHG inventory emissions claim or a project's GHG emission reduction claim.

These can be downloaded for a fee at <http://www.iso.org/>.

An additional source of general reporting requirements for greenhouse gases is the California Climate Action Registry (CCAR). CCAR has published several widely used general reporting and project reporting protocols. These can be found at <http://www.climateregistry.org>.

E.3 Additional Resources

Berlinkski, M.P. (2006). *Quantifying Emissions Reductions from New England Offshore Wind Energy Resources*. Thesis.

California Public Utilities Commission (1998). *Protocols and Procedures for the Verification of Costs, Benefits, and Shareholder Earnings from Demand-Side Management Programs*. Prepared by the California Demand Side Management Advisory Committee. <http://www.calmac.org/cadmac-protocols.asp#>

California Public Utilities Commission (2006). *Protocols for Estimating the Load Impacts from DR Program*. Draft Version 1. Prepared by Summit Blue Consulting, LLC, and Quantum Consulting, Inc. <http://www.cpuc.ca.gov/static/HotTopics/1energy/draftdrloadimpactprotocols.doc>

California State Governor's Office (2001). *California Standard Practice Manual: Economic Analysis of Demand-Side Management Programs*. <http://www.energy.ca.gov/greenbuilding/documents/background/07-J_CPUC_STANDARD_PRACTICE_MANUAL.pdf>

Chambers, A., D.M. Kline, L. Vimmerstedt, A. Diem, D. Dismukes, and D. Mesyanzhinov, D. (2005). *Comparison of Methods for Estimating the* NO_X *Emission Impacts of Energy Efficiency and Renewable Energy Projects: Shreveport, Louisiana Case Study*. NREL/TP-710-37721.

Con Edison (2007). *Demand Side Bidding Guidelines*. <http://www.coned.com/sales/business/targetedRFP 2007.asp>

Connecticut Department of Public Utility Control (2004). *Program Savings Documentation (PSD)*. Prepared as part of The Connecticut Light and Power Company's and The United Illuminating Company's Conservation and Load Management (C&LM) Plan for Year 2005, Docket 04-11-01. <http://www.state.ct.us/dpuc/ecmb/index.html>

Efficiency Vermont (2002). *Technical Reference User Manual (TRM)*. <http://www.efficiencyvermont.com>

Electric Power Research Institute (1991). *Impact Evaluation of Demand-Side Management Programs: Volume 1: A Guide to Current Practice*. <http://www.epri.com>

Electric Power Research Institute (1992). *DSM Evaluation—Six Steps for Assessing Programs*. <http://www.epri.com>

Electric Power Research Institute (2001). *Market Transformation: A Practical Guide to Designing and Evaluating Energy Efficient Programs*. <http://www.epri.com>

High, C., and K. Hathaway (2006). *Estimation of Avoided Emission Rates for Nitrogen Oxide Resulting from Renewable Electric Power Generation in the New England, New York and PJM Interconnection Power Market Areas*. Systems Group Inc.

International Petroleum Industry Environmental Conservation Association Climate Change Working Group and American Petroleum Institute (2007). *Oil and Natural Gas Industry Guidelines for Greenhouse Gas Reduction Projects*. <http://www.ipieca.org/activities/climate_change/climate_publications.php#17>

ISO New England (2004). *NEPOOL Marginal Emission Rate Analysis for the NEPOOL Environmental Planning Committee*.

Keith, G., D. White, and B. Biewald (2002). *The OTC Emission Reduction Workbook 2.1: Description and Users' Manual—Prepared for The Ozone Transport Commission*. Synapse Energy Economics.

Keith, G., and B. Biewald (2005). *Methods for Estimating Emissions Avoided by Renewable Energy and Energy Efficiency*. Prepared for the U.S. Environmental Protection Agency. Synapse Energy Economics

La Capra Associates and MSB Energy Associates (2003). *Electric Sector Emissions Displaced Due to Renewable Energy Projects in New England. February 2003 Analysis*. Prepared for Massachusetts Technology Collaborative.

Lawrence Berkeley National Laboratory (1999). *Guidelines for the Monitoring, Evaluation, Reporting, Verification, and Certification of Energy efficiency Projects for Climate Change Mitigation*. LBNL-41877. <http://ies.lbl.gov/iespubs/41877.pdf>

New Jersey Clean Energy Program (2004). *New Jersey Clean Energy Program Protocols to Measure Resource Savings*. <http://www.njcleanenergy.com/files/file/Protocols_REVISED_VERSION_1.pdf>

New York State Energy Research and Development Authority. Deemed Savings Database, Version 9.0. <http://www.nyserda.org>

Northwest Power and Conservation Council. Conservation Resource Comments Database. <http://www.nwcouncil.org/comments/default.asp>

Northwest Regional Technical Forum (RTF) documents. <http://www.nwcouncil.org/energy/rtf/Default.htm>

Pacific Consulting Services (1994). *Quality Assurance Guidelines for Statistical and Engineering Models*. 1994. Prepared for the California Demand Side Management Advisory Committee. <http://www.calmac.org/publications/2005.pdf>

Public Utility Commission of Texas (2003). *Deemed Savings, Installation & Efficiency Standards: Residential and Small Commercial Standard Offer Program, and Hard-to-Reach Standard Offer Program*. <http://www.puc.state.tx.us>

Public Utility Commission of Texas. 2005. *Measurement and Validation Guidelines*. <http://www.puc.state.tx.us/electric/projects/30331/052505/m%26v%5Fguide%5F052505.pdf>

Sebold, F., et al. (2001). *A Framework for Planning and Assessing Publicly Funded Energy Efficiency*. Prepared for Pacific Gas and Electric Company. <http://www.calmac.org/publications/20010301PGE0023ME.pdf>

United Nations Framework Convention on Climate Change (various years). *Methodologies for Clean Development Mechanism (CDM) Project Activities*. <http://cdm.unfccc.int/methodologies/index.html>

U.S. Department of Energy (2006). *Final Report on the Clean Energy/Air Quality Integration Initiative for the Mid-Atlantic Region*. <http://www.eere.energy.gov/wip/clean_energy_initiative.html>

U.S. Environmental Protection Agency (1995). *Conservation Verification Protocols: A Guidance Document for Electric Utilities Affected by the Acid Rain Program of the Clean Air Act Amendments of 1990*. SuDoc EP 4.8:C 76/3.

U.S. Environmental Protection Agency (2004). *Guidance on State Implementation Plan (SIP) Credits for Emission Reductions from Electric-Sector Energy Efficiency and Renewable Energy Measures*. <http://www.epa.gov/ttn/oarpg/t1/memoranda/ereseerem_gd.pdf>

U.S. Environmental Protection Agency (2007). *eGRID-Emissions and Generation Resource Integrated Database Web site*. <http://www.epa.gov/cleanenergy/egrid/index.html>

U.S. Environmental Protection Agency (2007). *Evaluation, Measurement, and Verification of Electricity Savings for Determining Emission Reductions from Energy Efficiency and Renewable Energy Actions*. <http://www.epa.gov/cleanenergy/pdf/ee-re_set-asides_vol3.pdf>

Vermont Energy Investment Corporation (2006). *Technical Reference Manual (TRM)*.

E.4 Program and Organization Web Sites

Building Owners and Managers Association (BOMA) International: <http://www.boma.org/TrainingAndEducation/BEEP/>

California's Appliance Efficiency Program (including California Title 20 Appliance Standards): <http://www.energy.ca.gov/appliances/index.html>

California Climate Action Registry: <http://www.climateregistry.org>

California Demand Response Programs: <http://www.energy.ca.gov/demandresponse/index.html>

California Energy Commission Efficiency Programs: <http://www.energy.ca.gov/efficiency/>

California Green Building Initiative: <http://www.energy.ca.gov/greenbuilding/index.html>

California Investor-Owned Utility Energy efficiency Programs: <http://www.californiaenergyefficiency.com/>

California Municipal Utilities Association: <http://www.cmua.org>

California Solar Initiative: <http://www.cpuc.ca.gov/static/energy/solar/index.htm>

The Climate Trust: <http://www.climatetrust.org>

Efficiency Vermont: <http://www.efficiencyvermont.com/pages/>

Efficiency Valuation Organization: <http://www.evo-world.org>

European Union Energy Efficiency Directive, measurement, monitoring, and evaluation Web site: <http://www.evaluate-energy-savings.eu/emeees/en/home/index.php>

International Energy Program Evaluation Conference: <http://www.iepec.org/>

Maine State Energy Program: <http://www.state.me.us/msep/>

Northeast Energy Efficiency Council: <http://www.neec.org>

Northeast Energy Efficiency Partnerships: <http://www.neep.org>

Northwest Energy Efficiency Alliance: <http://www.nwalliance.org/>

New York State Energy Research and Development Authority: <http://www.nyserda.org>

Texas Energy Efficiency Programs: <http://www.texasefficiency.com/>

Western Renewable Energy Generation Information System: <http://www.wregis.org/>

United Nations Framework Convention for Climate Change, Clean Development Mechanism: <http://cdm.unfccc.int/index.html>

U.S. Department of Energy:

- Efficiency and renewable energy: <http://www.eere.energy.gov>
- 1605b Program: <http://www.eia.doe.gov/environment.html>

U.S. Environmental Protection Agency:

- Clean Energy Programs: <http://www.epa.gov/solar/epaclean.htm>
- ENERGY STAR: <http://www.energystar.gov/>

World Business Council for Sustainable Development: <http://www.wbcsd.org>

World Resources Institute: <http://www.wri.org>

E.5 Notes

1. The Efficiency Valuation Organization (EVO). The IPMVP and related M&V resources can be found at <http://www.evo-world.org>.
2. Along with the FEMP M&V Guidelines, a number of other M&V resource documents, including some on the use of stipulations for determining savings, M&V checklists, and M&V resource lists, can be found at the Lawrence Berkeley National Laboratory Web site: <http://ateam.lbl.gov/mv/>.
3. The Guideline can be purchased at <http://www.ashrae.org>. As of the publication of this document, a new version of Guideline 14 is under development.

Appendix F: Renewables and Combined Heat and Power Project Measurement and Verification

This Guide addresses energy efficiency programs. However, other clean energy program types are related to efficiency. This appendix provides a brief overview of some of the approaches to the M&V of savings from renewable electrical energy projects and combined heat and power (CHP) projects.

F.1 Renewables Project Electricity Savings

This section introduces methods for determining savings from on-grid electric renewable energy projects and discusses some related issues. There are a variety of diverse technologies that convert renewable energy into electricity. Despite individual differences, these renewable energy technologies supply electricity and reduce the use of other grid-connected sources. In contrast, energy efficiency projects reduce electricity consumption. The implication is that renewable energy project M&V for electricity savings is simpler than energy efficiency M&V. This is because, in most instances, M&V simply involves measuring the electrical output of the subject system to determine the quantity of other grid-based electricity "saved." For renewable generation that produces emissions, however, a net emissions rate for each pollutant will be needed, adding a complication to the emissions estimation step. Life cycle emissions may also be important to compare in cases where major differences between renewables and baseline systems occur upstream.

The renewable energy projects covered in this chapter are the installation of devices or systems that displace grid electricity production through the use of renewable energy resources. Examples of renewable technologies include solar photovoltaics, biomass conversion systems (e.g., landfill gas methane recovery projects), and wind generators.

F.1.1 M&V Approaches and Options

There are two general approaches for calculating electricity savings:

1. **Direct measurement.** This approach assumes that the electricity produced by the renewable system displaces energy that would have been provided by an electric generating unit (EGU). With this one-for-one replacement approach, one only needs to directly measure the net amount of energy produced by the renewable system. This approach is most common with photovoltaic, wind, and biomass electricity production projects (assuming there is no supplementary firing with fossil fuels at the biomass facility).

2. **Net-energy use calculation.** With this approach, purchased electrical energy used at the project site during the reporting period is compared with a baseline to determine the savings in electricity purchases. When a baseline is adopted, there are four methods for calculating savings as defined in the 2003 IPMVP renewables protocol (IPMVP, 2003).

    - **Comparison with a control group.** Electricity consumption of the renewable energy system is compared with the electricity consumption of a control group, with similar characteristics under similar conditions. The control group is used as the baseline.

    - **Before and after comparison.** Electricity consumption of the renewable energy system is compared with the electricity consumption measured before the renewable system was installed for the same loads. The pre-installation situation is the baseline.

- **On and off comparison.** Electricity consumption with the renewable energy system "on" is compared to consumption with the system "off." The baseline equals the situation with the system "off."
- **Calculated reference method.** The baseline is determined with engineering calculations, and estimated electricity consumption is compared to metered energy use when the renewable energy system is in place. This approach has the weakness of using two different analyses methods (engineering estimates and metered data) to determine a difference—that is, the savings.

These four methods usually require measurement of electricity consumption or supply over an extended period in order to capture the variation due to changing climatic conditions.

The four IPMVP Options (A, B, C and D) can also be used for renewable energy projects. Options A and B involve measurements of system performance and are the most common. Option A involves stipulation of some parameters, while Option B requires maximum use of measurements in the energy savings analyses. Option C measures the change in whole-facility electricity use, usually with utility metering data, associated with the installation of the renewable system. Option D involves the use of computer simulations, calibrated with actual data, to determine savings from a renewable system installation.

F.1.2 Net Metering of Electrical Output and Fuel Use

In some situations, the electrical output of the renewable system is not directly indicative of electricity savings (and the avoided savings). These are when:

The system consumes electricity in order to produce electricity. The consumption is associated with what is known as *parasitic loads*. For example, a solar thermal electric system consumes electricity to power pumps that circulate fluid through the system. In these situations, either the parasitic loads have to be directly measured and subtracted from the measured output of the system, or a "net output" meter that accounts for parasitic loads is used.

The system consumes a fuel. An example is a landfill gas generation system that uses natural gas as a supplemental fuel. In these situations, incremental fuel usage must be accounted for when calculating energy savings.

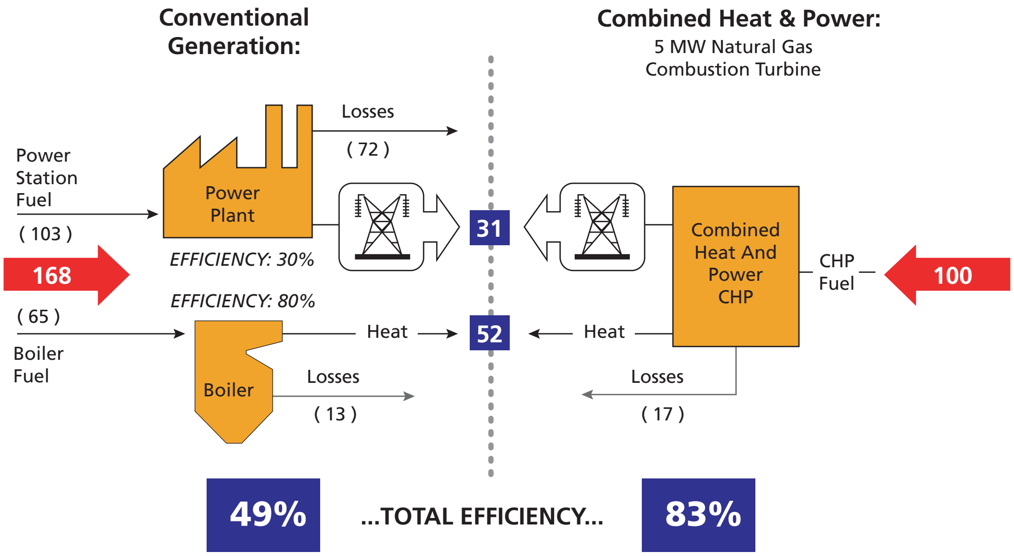
F.2 Efficiency Metrics for CHP Systems: Total System and Effective Electric Efficiencies[1]

CHP is an efficient approach to generating power and useful thermal energy from a single fuel source. CHP is used either to replace or supplement conventional separate heat and power (SHP) (e.g., central station electricity available via the grid and an onsite boiler or heater). Every CHP application involves the recovery of otherwise wasted thermal energy to produce additional power or useful thermal energy; as such, CHP offers energy efficiency and environmental advantages over SHP. CHP can be applied to a broad range of applications and the higher efficiencies result in lower emissions than SHP. The advantages of CHP broadly include the following:

- The simultaneous production of useful thermal energy and power in CHP systems leads to increased fuel efficiency.
- CHP units can be strategically located at the point of energy use. Such onsite generation prevents the transmission and distribution losses associated with electricity purchased via the grid from central station plants.
- CHP is versatile and can be designed for many different applications in the industrial, commercial and institutional sectors.

Figure F-1 shows how CHP can save energy compared to SHP.[2] CHP typically requires only two thirds to three quarters of the primary energy to produce the same

Figure F-1. CHP and SHP Energy Savings



Source: EPA, 2004.

thermal and electric service compared to separate heat and power. This reduced primary fuel consumption is key to the environmental benefits of CHP since burning the same fuel but using more of its energy means fewer emissions for the same level of output.

Efficiency is a prominent metric used to evaluate CHP performance and compare it to SHP. Two methodologies are most commonly used to determine the efficiency of a CHP system: total system efficiency and effective electric efficiency.

F.2.1 Key Terms Used in Calculating CHP Efficiency

Calculating a CHP system's efficiency requires an understanding of several key terms, described below.

- **CHP system.** The CHP system includes the unit in which fuel is consumed (e.g. turbine, boiler, engine), the electric generator, and the heat recovery unit that transforms otherwise wasted heat to useable thermal energy.

- **Total fuel energy input (Q_{FUEL}).** The energy associated with the total fuel input. Total fuel input is the sum of all the fuel used by the CHP system. The total fuel energy input is often determined by multiplying the quantity of fuel consumed by the heating value of the fuel.[3]

  Commonly accepted heating values for natural gas, coal, and diesel fuel are:

  - 1,020 Btu per cubic foot of natural gas
  - 10,157 Btu per pound of coal
  - 138,000 Btu per gallon of diesel fuel

- **Net useful power output (W_E).** Net useful power output is the gross power produced by the electric generator minus any parasitic electric losses. An example of a parasitic electric loss is the electricity that may be used to compress the natural gas before the gas can be fired in a turbine.

- **Net useful thermal output (SQ_{TH}).** Net useful thermal output is equal to the gross useful thermal output of the CHP system minus the thermal input. An example of thermal input is the energy of the condensate return and makeup water fed to a heat recovery steam generator. Net useful thermal output represents the otherwise wasted thermal energy that was recovered by the CHP system and used by the facility.

Gross useful thermal output is the thermal output of a CHP system utilized by the host facility. The term "utilized" is important here. Any thermal output that is not used should not be considered. Consider, for example, a CHP system that produces 10,000 pounds of steam per hour, with 90 percent of the steam used for space heating and the remaining 10 percent exhausted in a cooling tower. The energy content of 9,000 pounds of steam per hour is the gross useful thermal output.

F.2.2 Selecting a CHP Efficiency Metric

The selection of an efficiency metric depends on the purpose of calculating CHP efficiency.

- If the objective is to compare CHP system energy efficiency to the efficiency of a site's SHP options, then the *total system efficiency metric* may apply. Calculation of SHP efficiency is a weighted average (based on a CHP system's net useful power output and net useful thermal output) of the efficiencies of the SHP production components. The separate power production component is typically 33-percent-efficient grid power. The separate heat production component is typically a 75- to 85-percent-efficient boiler.
- If CHP electrical efficiency is needed for a comparison of CHP to conventional electricity production (i.e., the grid), then the *effective electric efficiency metric* may apply. Effective electric efficiency accounts for the multiple outputs of CHP and allows for a direct comparison of CHP and conventional electricity production by crediting that portion of the CHP system's fuel input allocated to thermal output.

Both the total system and effective electric efficiencies are valid metrics for evaluating CHP system efficiency. They both consider all the outputs of CHP systems and, when used properly, reflect the inherent advantages of CHP. However, since each metric measures a different performance characteristic, use of the two different metrics for a given CHP system produces different values.

For example, consider a gas turbine CHP system that produces steam for space heating with the following characteristics:

Fuel input (MMBtu/hr)	57
Electric output (MW)	5.0
Thermal output (MMBtu/hr)	25.6

According to the total system efficiency metric, the CHP system efficiency is 75 percent: (5.0 × 3.413 + 25.6) ÷ 57).

Using the effective electric efficiency metric, the CHP system efficiency is 68 percent: (5.0 × 3.413) ÷ (57 – (25.6 ÷ 0.8)).

Calculating Total System Efficiency

The most common way to determine a CHP system's efficiency is to calculate *total system efficiency*. Also known as *thermal efficiency*, the total system efficiency (η_o) of a CHP system is the sum of the net useful power output (W_E) and net useful thermal outputs (ΣQ_{TH}) divided by the total fuel input (Q_{FUEL}):

$$\eta_O = \frac{W_E + \Sigma Q_{TH}}{Q_{FUEL}}$$

The calculation of total system efficiency is a simple and useful method that compares what is produced (i.e., power and thermal output) to what is consumed (i.e., fuel). CHP systems with a relatively high net useful thermal output typically correspond to total system efficiencies in the range of 60 to 85 percent.

Note that this metric does not differentiate between the value of the power output and the thermal output; instead, it treats power output and thermal output as additive properties with the same relative value. In reality and in practice, thermal output and power output are not interchangeable because they cannot be converted easily from one to another. However, typical CHP applications usually have coincident power and thermal demands that must be met. It is reasonable, therefore, to consider the values of power and thermal output from a CHP system to be equal in many situations.

Calculating Effective Electric Efficiency

Effective electric efficiency calculations allow for a direct comparison of CHP to conventional power generation system performance (e.g., electricity produced from central stations, which is how the majority of electricity is produced in the United States). Effective electric efficiency accounts for the multiple outputs of CHP and allows for a direct comparison of CHP and conventional electricity production by crediting that portion of the CHP system's fuel input allocated to thermal output. The calculation of effective electric efficiency is analogous to the method many states use to apply a CHP thermal credit to output-based emissions estimates.

Effective electric efficiency (e_{EE}) can be calculated using the equation below, where (W_E) is the net useful power output, (SQ_{TH}) is the sum of the net useful thermal outputs, (Q_{FUEL}) is the total fuel input, and α equals the efficiency of the conventional technology that otherwise would be used to produce the useful thermal energy output if the CHP system did not exist:

$$\varepsilon_{EE} = \frac{W_E}{Q_{FUEL} - \sum (Q_{TH}/\alpha)}$$

For example, if a CHP system is natural gas–fired and produces steam, then α represents the efficiency of a conventional natural gas–fired boiler. Typical α values for boilers are 0.8 for a natural gas–fired boiler, 0.75 for a biomass-fired boiler, and 0.83 for a coal-fired boiler.

The effective electric efficiency is essentially the CHP net electric output divided by the fuel the CHP system consumes over and above what would have been used by conventional systems to produce the thermal output for the site. In other words, this metric measures how effectively the CHP system generates power once the thermal demand of a site has been met.

Typical effective electrical efficiencies for combustion turbine–based CHP systems are in the range of 50 to 75 percent. Typical effective electrical efficiencies for reciprocating engine–based CHP systems are in the range of 65 to 80 percent.

Obtaining the Required Data to Calculate CHP System Performance

Typically, CHP systems are sized so that their full electric and thermal output can be used during most of the year. Thermal output is always available from the CHP system when it is running; however, it is only useful when it can be applied to meet specific thermal loads at the site. The useful thermal output from the CHP system displaces load from a boiler, furnace, chiller, or other system. Many thermal loads, such as space heating, only occur for part of the year. As such, the utilization of the thermal output of a CHP system can vary with time of day, month, or season. The annual impact of these variations must be considered to accurately account for the efficiency benefits of CHP systems.

A reasonable M&V program for CHP systems must be able to credibly estimate the net power output and useful thermal output on an annual basis, yet impose only minimal additional burden on the end-user. An effective M&V plan must define the CHP system boundaries, identify applicable thermal loads and how they are served by the CHP system, include simple measurement and calculations approaches, and specify reporting requirements. The plan can be based on key performance assumptions and design estimates contained in initial permit applications. These assumptions can be verified with steady-state measurements at commissioning. However, the primary approach to verifying net power and useful thermal output of a system is long-term cumulative measurement or readings of power and thermal output from the system. These readings can be obtained through the installation of specific metering equipment (as an example, power metering is likely to be installed on most CHP systems; often, electric meters are required by an area's local utility as part of the interconnection requirements), or in many cases through the CHP system's automated control system, programmed to accumulate and log power and thermal data. Cumulative readings of system output can be collected either monthly or annually. The M&V plan should contain procedures to confirm the completeness of the information and the validity of any calculations that estimate thermal energy actually used based on measured system output.

The plan should also recognize that the CHP system may not operate for brief periods during the year due to planned maintenance or unscheduled outages. The availability[4] of CHP systems is an important component of overall system performance, and affects the reliability of power and thermal supply to the user. In general, the availability of CHP systems is high and the use of CHP systems operating in parallel to the grid often improves the reliability of energy supply to the site. The most recent comprehensive review of DG/CHP availability was conducted for Oak Ridge National Laboratory in 2003 (Energy and Environmental Analysis, 2004). Of the systems studied, the availability factor for reciprocating engines averaged 96 to 98 percent. Gas turbines had availability factors ranging from 93 to 97 percent.

F.3 Notes

1. This section was provided by the U.S. EPA.
2. Conventional power plant efficiency based on average U.S. fossil heat rate of 12,215 Btu/kWh (2004 eGRID) and average T&D losses of 7 percent; comparison assumes that thermal energy produced by the CHP system is used on site.
3. Fuel heating values are denoted as either lower heating value (LHV) or higher heating value (HHV). HHV includes the heat of condensation of the water vapor in the products of combustion. Unless otherwise noted, all heating value and efficiency measures in this section are reported on an HHV basis.
4. The availability factor is the proportion of hours per year that a unit "could run" (based on planned and unplanned maintenance) divided by the total hours in the year.

Appendix G: References

American Society of Heating, Refrigerating, and Air-Conditioning Engineers [ASHRAE] (2002). *Guideline 14 on Measurement of Demand and Energy Savings.*

Baumert, K., T. Herzog, and J. Pershing (2005). *Navigating the Numbers: Greenhouse Gases and International Climate Change Policy*. Washington, D.C. World Resources Institute.

California Public Utilities Commission [CPUC] (2004). *The California Evaluation Framework*. <http://www.calmac.org/publications/California_Evaluation_Framework_June_2004.pdf>

California Public Utilities Commission [CPUC] (2006). *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals*. <http://www.calmac.org/publications/EvaluatorsProtocols_Final_AdoptedviaRuling_06-19-2006.pdf>.

California State Governor's Office (2001). *California Standard Practice Manual: Economic Analysis of Demand-Side Management Programs*. <http://www.energy.ca.gov/greenbuilding/documents/background/07-J_CPUC_STANDARD_PRACTICE_MANUAL.pdf>

Efficiency Valuation Organization [EVO] (2007). *International Performance Measurement and Verification Protocol*. <http://www.evo-world.org>

Energy and Environmental Analysis, Inc. (2004). *Final Report: Distributed Generation Operational Reliability and Availability Database.* <http://www.eeainc.com/dgchp_reports/FinalReportORNLDGREL.pdf>

International Council for Local Environmental Initiatives [ICLEI] (2003). *Clean Air and Climate Protection Software.* <http://www.cacpsoftware.org/>

International Performance Measurement and Verification Protocol [IPMVP] (2003). *Concepts and Practices for Determining Energy Savings in Renewable Energy Technologies Applications, Volume III.* <http://www.evo-world.org>

ISO New England [ISO-NE] (2007). *Measurement and Verification of Demand Reduction Value from Demand Resources*. <http://www.iso-ne.com/rules_proceds/isone_mnls/index.html>

National Action Plan for Energy Efficiency (2006). *The National Action Plan for Energy Efficiency*. <http://www.epa.gov/cleanenergy/pdf/napee/napee_report.pdf>

National Action Plan for Energy Efficiency (2007a). *Guide to Resource Planning with Energy Efficiency*. Prepared by Snuller Price et al., Energy and Environmental Economics, Inc. <http://www.epa.gov/eeactionplan>

National Action Plan for Energy Efficiency (2007b). *Guide for Conducting Energy Efficiency Potential Studies*. Prepared by Philip Mosenthal and Jeffrey Loiter, Optional Energy, Inc. <http://www.epa.gov/actionplan>

New York State Energy Research and Development Authority [NYSERDA] (2006). *Non-Energy Impacts (NEI) Evaluation*. Prepared by Summit Blue Consulting, LLC, and Quantec, LLC.

Northeast Energy Efficiency Partnership [NEEP] (2006). *The Need for and Approaches to Developing Common Protocols to Measure, Track, and Report Energy Efficiency Savings in the Northeast*. Prepared by J. Michals and E. Titus. ACEEE Summer Study 2006.

Ozone Transport Commission [OTC] (2002). *OTC Emission Reduction Workbook (Version 2.1).* Prepared by Geoff Keith, David White, and Bruce Biewald.

PA Consulting Group (2006). *2005 Commercial and Industrial Programs Free-Ridership and Spillover Study Revised Report*.

Schiller Consulting (2007). *Survey of Energy Efficiency Evaluation Measurement and Verification (EM&V) Guidelines and Protocols.*

State of Wisconsin, Public Service Commission of Wisconsin (2007). *Focus on Energy Evaluation: Semiannual Report*. PA Government Services, Inc. <http://www.focusonenergy.com>

TecMarket Works (2002). *Non-Energy Benefits Cross-Cutting Report. Year 1 Efforts*. Prepared by J. Riggert, N. Hall, and T. Talerico.

TecMarket Works (2003). *Non-Energy Benefits to Implementing Partners from the Wisconsin Focus on Energy Program*. Prepared by N. Hall and J. Roth.

TecMarket Works (2005). *A Qualitative Program Effects Status Study*. Prepared by N. Hall.

U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy [DOE] (2003). *Program Management Guide: A Reference Manual for Program Management*. <http://www1.eere.energy.gov/ba/prog_mgmt_guide.html>

U.S. Department of Energy [DOE] (2006). *Final Report on the Clean Energy/Air Quality Integration Initiative for the Mid-Atlantic Region* <http://www.eere.energy.gov/wip/clean_energy_initiative.html>

U.S. Department of Energy [DOE] (2006). *109th Congress, Report on Section 1252 of the Energy Policy Act of 2005*.

U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy [DOE] (2007). *Impact Evaluation Framework for Technology Deployment Programs*. Prepared by J. Reed, G. Jordan, and E. Vine. <http://www.eere.energy.gov/ba/pba/km_portal/docs/pdf/2007/impact_framework_tech_deploy_2007_main.pdf>

U.S. Environmental Protection Agency [EPA] (1995). Conservation Verification Protocol.

U.S. Environmental Protection Agency [EPA] (2004). *Output-Based Regulations: A Handbook for Air Regulators.* <http://www.epa.gov/chp/documents/obr_final_9105.pdf>

U.S. Environmental Protection Agency [EPA] (2007). *eGRID—Emissions and Generation Resource Integrated Database Web site*. <http://www.epa.gov/cleanenergy/egrid/index.html>

World Resources Institute [WRI] and World Business Council for Sustainable Development [WBCSD] (2005a). *Measurement and Estimation Uncertainty for GHG Emissions*.

World Resources Institute [WRI] and World Business Council for Sustainable Development [WBCSD] (2005b). *GHG Protocol for Project Accounting*.

Funding and printing for this report was provided by the U.S. Department of Energy and U.S. Environmental Protection Agency in their capacity as co-sponsors for the National Action Plan for Energy Efficiency.

♲ Recycled/Recyclable—Printed with Vegetable Oil Based Inks on 100% (Minimum 50% Postconsumer) Recycled Paper