DOE/EE-0829

SEE Action
STATE & LOCAL ENERGY EFFICIENCY ACTION NETWORK

Energy Efficiency Program Impact Evaluation Guide

Evaluation, Measurement, and Verification Working Group

December 2012

The State and Local Energy Efficiency Action Network is a state and local effort facilitated by the federal government that helps states, utilities, and other local stakeholders take energy efficiency to scale and achieve all cost-effective energy efficiency by 2020.

Learn more at www.seeaction.energy.gov

ABOUT THIS DOCUMENT

This *Energy Efficiency Program Impact Evaluation Guide* describes the common terminology, structures, and approaches used for determining (evaluating) energy and demand savings as well as avoided emissions and other non-energy benefits resulting from facility (non-transportation) energy efficiency programs that are implemented by local governments, states, utilities, private companies, and nonprofits. While this guide does not recommend specific approaches, it provides context, planning guidance, and discussion of issues that determine the most appropriate evaluation objectives and best practices approaches for different efficiency portfolios. By using standard evaluation terminology and structures and best practices approaches, evaluations can support the adoption, continuation, and expansion of effective efficiency actions.

The primary audiences for this guide are energy regulators; public and private energy efficiency portfolio administrators such as utilities, nonprofit organizations, and government agencies; program implementers; and evaluators looking for guidance on the following:

- The evaluation process and approaches for determining program impacts
- Planning evaluation efforts
- Key issues associated with establishing evaluation frameworks for improving the efficacy of energy efficiency portfolios, documenting the impacts of such portfolios, and comparing demand- and supply-side resources.

Introductory portions and appendices are also intended for policymakers seeking general information about efficiency program impact evaluation as well as the basic principles of process and market evaluations and cost-effectiveness analyses. Although the guide is not directly intended for expert evaluation practitioners who can rely on more detailed and specific resources that are referred to in this guide, it offers introductions to and summaries of evaluation topics that can be useful for explaining concepts and standard practices to clients, new staff, stakeholders, and others who could benefit from a refresher on principles.

This 2012 version of the guide is an update to the 2007 National Action Plan for Energy Efficiency *Model Energy Efficiency Program Impact Evaluation Guide*. Prepared by Steve Schiller, Schiller Consulting, Inc. www.epa.gov/eeactionplan.

The *Energy Efficiency Program Impact Evaluation Guide* was developed as a product of the State and Local Energy Efficiency (EPA) Action Network (SEE Action), which is facilitated by the U.S. Department of Energy and the U.S. Environmental Protection Agency. Content does not imply an endorsement by the individuals or organizations that are part of SEE Action working groups or reflect the views, policies, or otherwise of the federal government.

This effort was funded by the Permitting, Siting, and Analysis Division of the U.S. Department of Energy's Office of Electricity Delivery and Energy Reliability under Lawrence Berkeley National Laboratory Contract No. DE-AC02-05CH11231.

If this document is referenced, it should be cited as:
State and Local Energy Efficiency Action Network. 2012. *Energy Efficiency Program Impact Evaluation Guide*. Prepared by Steven R. Schiller, Schiller Consulting, Inc., www.seeaction.energy.gov.

FOR MORE INFORMATION

For more information regarding the *Energy Efficiency Program Impact Evaluation Guide*, please contact:

Michael Li
U.S. Department of Energy
Email: Michael.Li@ee.doe.gov

Carla Frisch
U.S. Department of Energy
Email: Carla.Frisch@ee.doe.gov

Regarding the State and Local Energy Efficiency Action Network, please contact:

Johanna Zetterberg
U.S. Department of Energy
Email: Johanna.Zetterberg@ee.doe.gov

Acknowledgments

The following people provided key input to the development of this version of the guide by providing materials, sidebars, and/or extremely valuable review and input on draft versions (*denotes member of SEE Action's EM&V Working Group):

- Jeffrey Brown, Robyn DeYoung, and Nikolaas Dietsch, U.S. Environmental Protection Agency
- Tom Eckman, Northwest Power Planning Council
- Hilary Forster, Consortium for Energy Efficiency*
- Donald Gilligan, National Association of Energy Services Companies*
- Fred Gordon, Energy Trust of Oregon*
- Dennis Hartline, Maryland Energy Administration*
- M. Sami Khawaja, Hossein Haeri, Tina Jayaweera, and David Sumi, The Cadmus Group
- Stephen Kromer, Kromer Engineering
- Jennifer Meissner, New York State Energy Research and Development Authority*
- Mike Messenger, Itron
- Julie Michals* and Elizabeth Titus, Northeast Energy Efficiency Partnerships
- Peter Miller, Natural Resources Defense Council*
- Jane Peters, Research Into Action
- Mitch Rosenberg, DNV KEMA
- Lisa Skumatz, Skumatz Economic Research Associates
- Rodney Sobin, Alliance to Save Energy*
- Annika Todd, Lawrence Berkeley National Laboratory
- Carol Zabin, University of California, Berkeley.

Dr. Khawaja provided substantial input into several sections of this guide, as did the EPA staff on the Avoided Emissions Chapter; their input is therefore particularly acknowledged.

Mark Wilson provided editing services. Publication and graphics were provided by the National Renewable Energy Laboratory.

List of Acronyms

A

ACEEE: American Council for an Energy-Efficient Economy

AEA: American Evaluation Association

ANSI: American National Standards Institute

ASHRAE: American Society of Heating, Refrigerating and Air-Conditioning Engineers

B

BAU: business as usual

BM: build margin (for electric generating units)

Btu: British thermal units

C

CAIR: Clean Air Interstate Rule

CALMAC: California Measurement Advisory Council

C&S: (efficiency) codes and standards

CDD: cooling degree day

CDM: (United Nations Framework Convention on Climate Change) Clean Development Mechanism

CEMS: Continuous Emission Monitoring Systems

CFL: compact fluorescent light bulb

CMVP: Certified Measurement and Verification Professionals

CO_2: carbon dioxide

CPUC: California Public Utility Commission

CSA: conditional savings analysis

CV: contingent valuation

Cx: commissioning

D

DEER: Database for Energy Efficiency Resources (California)

DOE: U.S. Department of Energy

DR: demand response

DSM: demand-side management

E

E&T: education and training

ECM: energy conservation measure

EE: energy efficiency

EERS: energy efficiency resource standard

EGU: electric generating unit (a power plant)

EM&V: evaluation, measurement, and verification

EPA: U.S. Environmental Protection Agency

ER: emission rate

ESCO: energy services company

ETO: Energy Trust of Oregon

EUL: effective useful life

F

FEMP: Federal Energy Management Program

FERC: Federal Energy Regulatory Commission

G

GHG: greenhouse gas

H

HDD: heating degree day

HERS: Home Energy Rating System

HHV: higher heating value

HVAC: heating, ventilation, and air conditioning

I

IPCC: Intergovernmental Panel on Climate Change

IPM: integrated planning model

IPMVP: International Performance Measurement and Verification Protocol

IRP: integrated resource planning

ISO: independent system operator or International Organization for Standardization

K

kW: kilowatt

kWh: kilowatt-hour

L

lb: pound

M

M&V: measurement and verification

MARAMA: Mid-Atlantic Regional Air Management Association

MMBtu: million Btu

MT: market transformation

MW: megawatt

MWh: megawatt-hour

N

NAESCO: National Association of Energy Service Companies

NEB: non-energy benefit

NEEA: Northwest Energy Efficiency Alliance

NEI: non-energy impact

NERC: North American Reliability Corporation

NOMAD: naturally occurring market adoption rate

NO_x: nitrogen oxide

NPV: net present value

NTG: net-to-gross

NTGR: net-to-gross ratio

O

O&M: operations and maintenance

OM: operating margin (for electric generating units)

P

PACT: program administrator cost test

PCT: participant cost test

PMP: Performance Measurement Protocols for Commercial Buildings (ASHRAE)

PSC: Public Service Commission

PUC: Public Utilities Commission

Q

QAG: quality assurance guideline

QA/QC: quality assurance/quality control

QEM: quasi-experimental methods

R

RASS: residential appliance saturation studies

RCx: retro-commissioning

RCT: randomized controlled trial

RD&D: research, development, and demonstration

REED: Regional Energy Efficiency Database

RFP: request for proposal

RFQ: request for qualifications

RGGI: Regional Greenhouse Gas Initiative

RIM: ratepayer impact test

RTF: Regional Technical Forum

S

SCT: societal cost test

SIP: State (air pollution reduction) Implementation Plan

SO_2: sulfur dioxide

SPM: (California) Standard Practice Manual

SPT: standardized project tracking

T

TBE: theory-based evaluation

T&D: transmission and distribution

TRC: total resource cost test

TRM: technical reference manual

U

UMP: Uniform Methods Project

V

VCS: Verified Carbon Standard

W

WBCSD: World Business Council for Sustainable Development

WRI: World Resources Institute

Table of Contents

List of Tables

List of Figures

Executive Summary

ES.1 TOPICS COVERED BY THIS GUIDE AND INTENDED AUDIENCES

This *Energy Efficiency Program Impact Evaluation Guide* describes and provides guidance on approaches for determining and documenting energy and non-energy benefits resulting from end-use energy efficiency programs and portfolios of programs. It specifically focuses on impact evaluations for programs designed to reduce facility (e.g., home, commercial building, factory) energy consumption and/or demand as well as related air emissions. This guide's objective is to support the implementation of effective energy efficiency actions by providing information on standard procedures and best practices for planning and conducting evaluations and reporting results. To this end, this guide accomplishes the following:

- Defines a systematic evaluation planning and implementation process
- Describes several standard approaches for determining energy and demand savings (as well as avoided emissions and other non-energy impacts)
- Defines key terms related to energy efficiency evaluation
- Provides guidance on key evaluation issues
- Lists publicly available energy efficiency evaluation resources.

The programs primarily addressed in this guide are voluntary; that is, program participants choose to take the efficiency actions as a result of some form of inducement. This guide does not focus on, but does touch on, evaluating mandatory requirements for efficiency such as found in codes and standards. Similarly, the guide only briefly addresses evaluating programs for which energy savings are an indirect benefit, such as contractor training programs.

The audiences for this guide are program designers, implementers, administrators, evaluators, and public agency officials who oversee and implement energy efficiency programs. Introductory portions of this guide are intended for policymakers seeking information about the basic principles of impact evaluation. Those looking for just the basics may want to read only through Chapter 3 and refer to the appendices for overviews of other evaluation types, definitions, and references. Some readers who are new to evaluation assignments can benefit from reading the entire document, while others may benefit from focusing on the evaluation issues and planning chapters (Chapters 7 and 8, respectively) and using the rest of the document as a reference. Although the guide is not intended for expert evaluation practitioners, they may find it useful for explaining evaluation concepts to those without their expertise.

Documents from the U.S. Department of Energy's (DOE's) Uniform Methods Project (UMP) serve as a companion set to this guide and include model evaluation plans for specific energy efficiency measures and program categories (e.g., residential lighting, refrigerators, commercial cooling).[1]

ES.2 DEFINITION AND IMPORTANCE OF ENERGY EFFICIENCY EVALUATION

Energy efficiency evaluation includes any of a range of assessment studies and other activities aimed at determining the effects of an energy efficiency program. Evaluations can document program performance, operations, changes in energy efficiency markets, and cost-effectiveness. There are three broad categories of efficiency program evaluations: impact evaluations, process evaluations, and market evaluations. Although this guide focuses on impact evaluations, it is helpful to know the purposes and goals of all three:

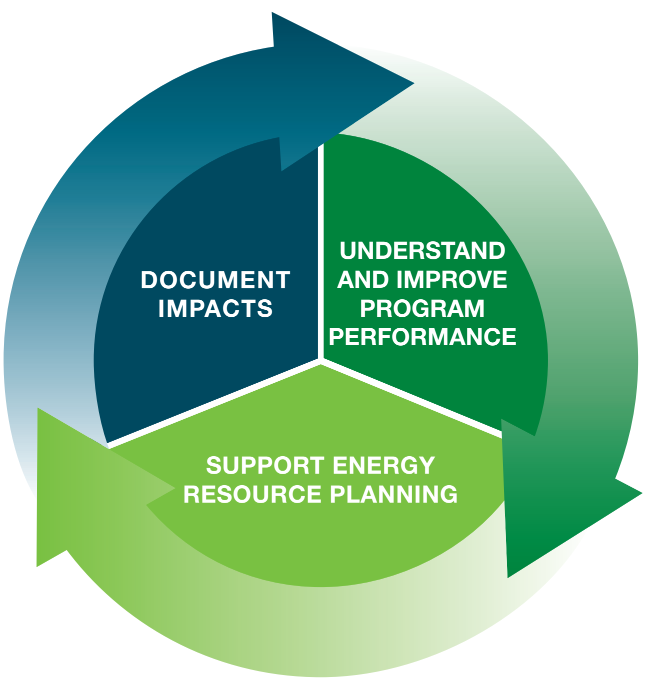
- **Impact evaluations:** assessments that determine and document the direct and indirect benefits of an energy efficiency program. Impact evaluation involves real-time and/or retrospective assessments of the performance and implementation of an efficiency program or portfolio of programs. Program benefits, or impacts, can include energy and demand savings and non-energy benefits (sometimes called co-benefits, with examples being avoided emissions, health benefits, job creation and local economic development, energy security, transmission and distribution benefits, and water savings). Impact evaluations also support cost-effectiveness analyses aimed at identifying relative program costs and benefits of energy efficiency as compared to other energy resources, including both demand- and supply-side options.
- **Process evaluations:** formative, systematic assessments of an energy efficiency program. They document program operations and identify and recommend improvements that are likely to increase the program's efficiency or effectiveness for acquiring energy efficiency resources, preferably while maintaining high levels of participant satisfaction.
- **Market evaluations:** assessments of structure or functioning of a market, the behavior of market participants, and/or market changes that result from one or more program efforts. Market evaluation studies may include estimates of the current market role of energy efficiency (market baselines), as well as the potential role of efficiency in a local, state, regional, or national market (potential studies). Market evaluation studies indicate how the overall supply chain and market for energy efficiency products works and how they have been affected by a program(s).

Market evaluations are critical for, but not exclusively used for, programs with market transformation elements and objectives. Examples of market evaluations are potential studies, baselines studies, and market effects studies.

Evaluations have three primary objectives, as shown in Figure ES.1:

- **Document the benefits** (i.e., impacts) of a program and determine whether the subject program (or portfolio of programs) met its goals
- **Identify ways to improve current and future programs** through determining why program-induced impacts occurred
- **Support energy demand forecasting and resource planning** by understanding the historical and future resource contributions of energy efficiency as compared to other energy resources.

FIGURE ES.1: Evaluation objectives



EVALUATION SUPPORTS SUCCESSFUL EFFICIENCY PROGRAMS

Documenting the benefits of efficiency, using credible and transparent methods, is a key component of successfully implementing and expanding the role of efficiency in providing secure, stable, reliable, clean, and reasonably priced energy. Therefore, evaluation is not an end unto itself but an effective tool for supporting the adoption, continuation, and expansion of energy efficiency programs, and thus the efficient use of energy.

Many energy efficiency evaluations are oriented toward developing retrospective estimates of energy savings attributable to a program to demonstrate in regulatory proceedings that public or energy consumer funds were properly and effectively spent. Beyond documenting savings and attribution, though, is the role of evaluation in improving programs and providing a basis for future savings estimates in resource plans. Therefore, evaluation both fosters more effective programs and justifies increased levels of investment in energy efficiency as a long-term, reliable energy resource. Perhaps the imperative for conducting evaluation is best described by a quote attributed to John Kenneth Galbraith: "Things that are measured tend to improve."[2]

ES.3 IMPACT EVALUATION METRICS

One or more of the following three metrics are usually reported as the output of impact evaluations:

- **Estimates of gross (energy and/or demand) savings.** These are the changes in energy consumption and/or demand that result directly from program-related actions taken by participants in an efficiency program, regardless of why they participated.
- **Estimates of net (energy and/or demand) savings.** These are the changes in energy consumption or demand that are attributable to an energy efficiency program. 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, to one degree or another, without the program now or in the near future) and participant and non-participant spillover (i.e., savings that result from actions taken as a result of a programs's influence but which are not directly subsidized or required by the program). Net savings may also include consideration of *market effects* (changes in the structure of a market).

  Determining net savings involves separating out the impacts that are a result of influences other than the program being evaluated, such as consumer self-motivation or effects of prior and/or other programs. Given the range of influences on consumers' energy consumption and the complexity in separating out both short-term and long-term market effects caused by the subject programs (and other programs), attributing changes to one cause (i.e., a particular program) or another can be quite complex. This is compounded by a general lack of consensus among policymakers and regulators on which short-term and long-term market influences and effects should be considered when determining net savings. Net savings are discussed in Chapter 5.

- **Estimates of non-energy benefits (NEBs).** These are the impacts associated with program implementation or participation aside from energy and demand savings. These results can be positive or negative. Some examples include reduced emissions and environmental benefits, productivity improvements, jobs created and local economic development, reduced utility customer disconnects, greater comfort for building occupants, lower maintenance costs due to better equipment, or increased maintenance costs due to new and more complex systems. NEBs are discussed in Section 7.9.

ES.4 ENERGY SAVINGS ESTIMATES AND UNCERTAINTY: "HOW GOOD IS GOOD ENOUGH?"

Each of the bullets in Section ES.3 above defines an "estimate" versus an exact value. This is because energy and demand savings as well as non-energy benefits resulting from efficiency actions cannot be directly measured. Instead, savings and benefits are based on counterfactual assumptions. Using counterfactual assumptions implies that savings are estimated to varying degrees of accuracy by comparing the situation (e.g., energy consumption) after a program is implemented (the reporting period) to what is assumed to have been the situation in the absence of the program (the "counterfactual" scenario, known as the baseline). For energy impacts, the baseline and reporting period energy use are compared, while controlling (making adjustments) for factors unrelated to energy efficiency actions, such as weather or building occupancy. These adjustments are a major part of the evaluation process; how they are determined can vary from one program type to another and from one evaluation approach to another.

Because the indicated values are estimates, their use as a basis for decision making can be challenged if their sources and level of accuracy are not described. Therefore, evaluation results, like any estimate, should be reported as "expected values"; that is, based on the impact evaluation, values are expected to be correct within an associated level of certainty. Minimizing uncertainty and balancing evaluation costs with the value of the indicated evaluation information are at the heart of the evaluation process and leads to perhaps the most fundamental evaluation question: "How good is good enough?" This question is a short version of asking (1) what level of certainty is required for energy savings estimates resulting from evaluation activities, and (2) is that level of certainty properly balanced against the amount of effort (e.g., resources, time, money) used to obtain that level of certainty?

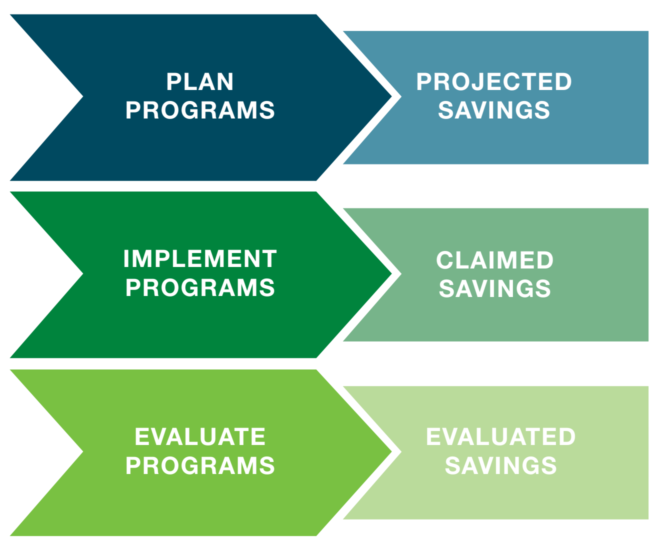
Two principles are important when considering "how good is good enough": (1) energy efficiency investments should be cost effective, and (2) evaluation investments should consider risk management principles and thus balance the costs of evaluation against the value of the information derived from evaluation (i.e., evaluation should also be cost effective). The value of the information is directly related to the risks of underestimating or overestimating the benefits (savings) and costs associated with efficiency investments. These risks might be associated with errors of commission or errors of omission. An error of commission might be overestimating savings, which in turn can result in continuing programs that are not cost effective and/or overpaying contractors, administrators, and participants. An error of omission, on the other hand, might be associated with underestimating savings or not implementing efficiency actions because of the difficulty in documenting savings, both of which can result in underinvesting in efficiency and relying on other energy resources that have their own risks and uncertainties, such as fuel costs and environmental impacts.

ES.5 EVALUATION AND THE EFFICIENCY PROGRAM PROCESS

As shown in Figure ES.2, the efficiency program process consists of planning, implementing, and evaluating activities. Throughout this process, savings values are typically indicated based on estimates prepared as part of each activity. One way to describe these savings is with the following classifications, also displayed in Figure ES.2:

- **Projected savings:** values reported by a program implementer or administrator before the efficiency activities are completed

FIGURE ES.2: Workflow and reporting for planning, implementing, and evaluating efficiency programs



- **Claimed savings:** values reported by a program implementer or administrator after the efficiency activities have been completed
- **Evaluated savings:** values reported by an independent third-party evaluator after the efficiency activities and impact evaluation have been completed. The designation of "independent" and "third-party" is determined by those entities involved in the use of the evaluations and may include evaluators retained, for example, by the program administrator or a regulator.

With respect to the evaluation activities, they can also be described as consisting of three phases: planning, implementation, and reporting, as shown in Figure ES.3 and described in the next subsections.

ES.5.1 Planning Impact Evaluations

The following provide the basic steps in planning impact evaluations:

1. Define the evaluation objectives and metrics in the context of the evaluated program's (or portfolio's) intended benefits, risks, and policy objectives.
2. Select appropriate evaluation approach(es) and prepare a program evaluation plan that takes into account the critical evaluation issues and the expectation for reliability (certainty) of evaluated impacts.
3. Define data collection requirements.

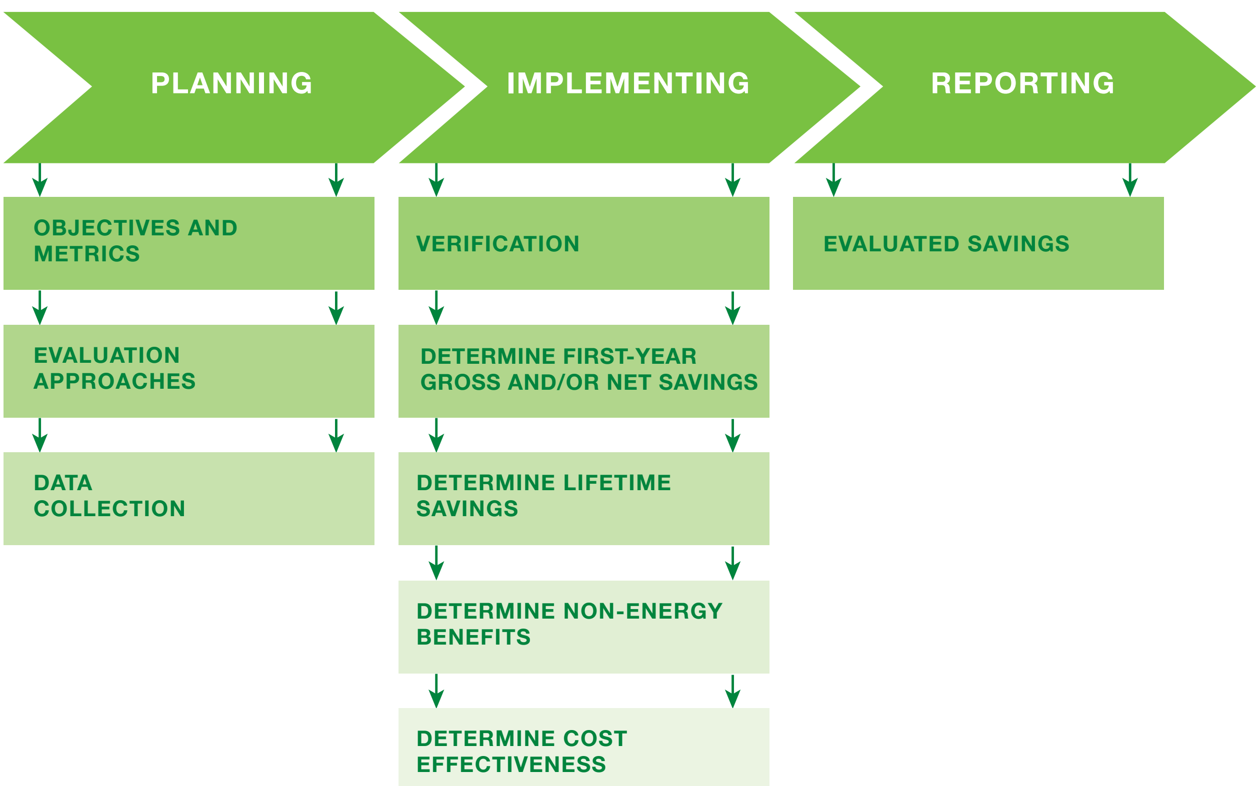
ES.5.2 Implementing Impact Evaluations

The impact evaluation is conducted through the following steps:

1. Verify actual implementation of the program, for example, by confirming installation and proper operation of the energy efficiency measures. This usually also includes auditing and validating assumptions used in the program planning process and checking program tracking databases, project applications, and other documentation and related data records for accurate recording of information.
2. Determine first-year program energy (and demand) savings using one of the following approaches (which are further defined and described in Chapters 3 and 4):

   a. Measurement and verification (M&V): a project-by-project approach involving estimating energy and/or demand savings by determining the savings for a representative sample of projects and applying these projects' savings to the entire population (i.e., the program). Options for conducting M&V are defined in the International Performance Measurement and Verification Protocol (IPMVP) and include two end-use metering options, billing regression analysis, and computer simulation. This approach determines gross savings values; net savings can be determined with program-wide adjustments to the gross savings values.

FIGURE ES.3: Evaluation activities workflow



b. Deemed savings values: stipulations based on historical and verified data (in some cases using the results of prior M&V studies). Similarly, *deemed savings calculations* are standardized algorithms. Both deemed savings values and deemed savings calculations should only be used with well-defined energy efficiency measures that have documented and consistent savings values. This approach determines gross savings values or net savings values, if net-to-gross ratios are included in the deemed savings values or calculations.

c. Large-scale consumption data analysis: uses metered energy use data to compare the energy use of the program participants with the energy use of a control group. The control group can be either program nonparticipants, as is the case with randomized controlled trials, or participants, as is the case with some quasi-experimental methods. If the program participants are used, their energy use before the program and after the program are compared; in effect, this means that each participant is his/her own non random control group. All of these methods can provide results that are either gross or net savings values.

In some cases, the three approaches listed above are combined, particularly the deemed savings and M&V approaches. Portfolios of programs also often use different approaches for different programs to determine total portfolio savings. Multiple-year programs may also conduct detailed measurement-based studies (e.g., M&V) for one year of the program and then apply the savings values (deemed savings) for other program years.

3. Convert, as needed, first-year gross program energy (and demand) savings to first-year net program savings using a range of possible considerations as discussed in Chapters 3 and 5.
4. Determine lifetime savings, which are the expected energy (and demand) savings over the lifetime of the measures that are implemented in the efficiency program. These savings are usually calculated by multiplying the first-year annual energy use reduction associated with the subject measures by the expected life of these measures with possible consideration of factors such as performance degradation or in some cases consideration of rebound (an increased level of service that is accompanied by an increase in energy use as a result of a program). Section 7.3 discusses savings persistence.
5. Determine non-energy benefits (NEBs) using a range of subjective and objective analytical tools. Determining avoided emissions, which is the primary NEB addressed in this guide, is discussed in Chapter 6. Evaluating other NEBs is discussed in Section 7.9.
6. Determine the program's cost-effectiveness using one or more of the common cost-effectiveness tests. Inputs into these tests are the lifecycle net or gross energy and demand savings and possibly one or more non-energy benefits. See Appendix B for an overview of cost-effectiveness analyses.

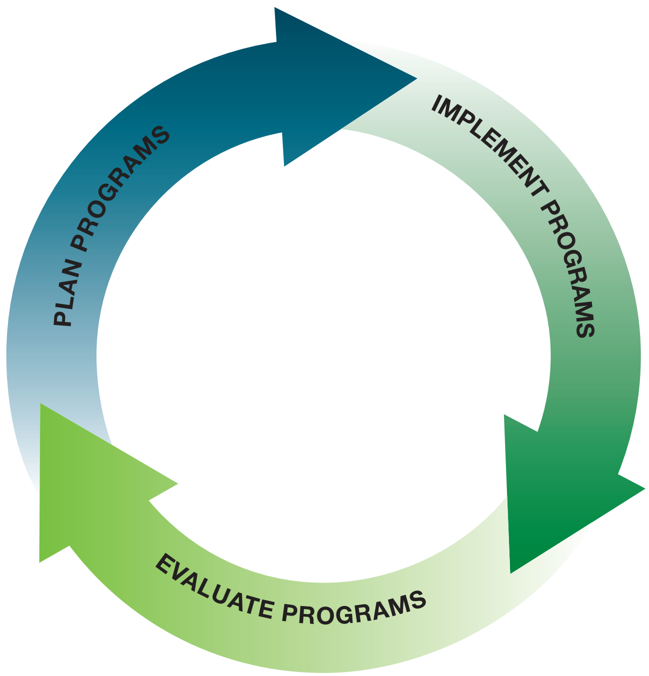
The evaluation approaches described in this guide are often referred to as "bottom-up" approaches because they add up the savings from measures and projects to determine program impacts, and they add up the impacts of programs to determine total portfolio impacts. Another evaluation category, called "top-down," uses approaches that rely on energy consumption data or per-unit energy consumption indicators (e.g., energy consumption per-unit of output or per person) defined by market sector, utility service territory, or a geographic region (e.g., a state or region). Top-down evaluation is not commonly used for evaluation of efficiency programs and portfolios, although interest in the approach is growing, and it has advantages over bottom-up evaluations. A section of Appendix B covers top-down evaluation.

ES.5.3 Evaluation Planning Characteristics and Frameworks

The following are best practice characteristics for evaluations:

- Evaluation is integral to a typical cyclic planning-implementation-evaluation process. Therefore, evaluation planning is part of the program planning process, including the alignment of implementation and evaluation budgets and schedules. This is done so that evaluation efforts can support program implementation and provide timely evaluation results for improving existing programs and informing future program and energy resource planning. See Figure ES.4.
- The evaluation process is designed to support the policy goals of the energy efficiency programs being evaluated by providing appropriate documentation of progress toward the goals, as well as feedback required by program administrators and implementers to continuously improve the programs and plan future efforts.
- Evaluation budgets and resources are adequate to support, over the entire evaluation, the evaluation goals and the level of quality (certainty) expected in the evaluation results. Reported values for metrics are those that are "most likely" and not biased to be overly conservative or overly aggressive.
- 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.

FIGURE ES.4: Evaluation is integral to a typical cyclic planning-implementation-evaluation process



- Evaluations are complete, readable, fair, accurate, transparently documented, relevant, and actionable, as well as balanced between certainty of results and costs to achieve the results. They also follow the American Evaluation Association's guiding principles, which are listed in Section 7.6.

With the above characteristics in mind, individual entities can define their own policy-specific program evaluation requirements. Jurisdictions such as states can establish and document their evaluation requirements in a hierarchy of documents. A useful structure of planning documents includes the following (see Figure ES.5):

- **Evaluation framework.** A framework is a primary document that lays out evaluation principles, metrics, allowable approaches, definitions, and metrics for determination of gross and/or net savings, reporting requirements, schedules, and the roles and responsibilities of various entities. An evaluation framework document tends to be "fixed" for several years, but of course can be updated periodically. It often sets the expectations for the content and scope of the other evaluation documents. This is perhaps the principle document that all stakeholders can focus on and provide high-level input to—the "forest versus the trees" of evaluation planning.
- **Portfolio cycle EM&V plan.** This plan indicates the major evaluation activities that will be conducted during the evaluation cycle (typically one, two, or three years). It includes the budget and allocation among the programs, measures, and market sectors, as applicable.
- **Evaluation activity-specific detailed plans.** Evaluation plans are created for each of the major evaluation activities (typically the evaluation of an energy efficiency program but may include studies such as market assessments) in a given cycle prior to the time each activity is launched.
- **Project-specific plans.** Project-specific plans may be required for custom project sites that are analyzed and inspected.

Also complementary to this hierarchy of planning documents is a reporting structure that can include individual site evaluation reports, program reports, and annual portfolio reports.

Another typical resource document for large-scale efficiency portfolios (such as those for a state or regional consumer-funded efficiency program) is a technical reference manual (TRM). A TRM is a database of standardized, state- or region-specific deemed savings calculations and associated deemed savings values for well-documented energy efficiency measures. Energy efficiency program administrators and implementation contractors use TRMs to reduce evaluation costs and uncertainty.

ES.5.4 Evaluation Planning Issues

The evaluation requirements described in each of the planning documents listed above are determined by the program objectives, regulatory mandates (if any), expectations for quality (i.e., reliability) of the evaluation results, available budgets, timing of reporting deadlines, intended uses of the evaluation results, and other factors that can vary across jurisdictions and programs. In this guide (Chapter 8), 14 key evaluation planning issues are presented and discussed to help define policy-specific program evaluation requirements:

1. What are the policy and/or regulatory goals that are the basis for the efficiency programs, and what are the evaluation objectives, metrics, and research issues that support the program policies and/or regulations?
2. What are the evaluation principles that drive the effort?
3. What is the scale and budget of the evaluation effort?
4. Who will conduct the evaluations, how is an independent evaluation defined, and what are the relative EM&V roles between implementers, evaluators, regulators, stakeholders, and others?

5. Is performance determined on the basis of net or gross savings? What factors are included in defining net savings?
6. What are the baselines against which savings are determined?
7. What is the reporting "boundary"? Are transmission and distribution (T&D) losses included, and how "granular" will the results be?
8. What are the schedules for implementing the evaluation and reporting?
9. What impact evaluation approaches will be used?
10. What are expectations for savings determination certainty (confidence and precision)?
11. Which cost-effectiveness tests will be used?
12. How are evaluated savings estimates applied—looking back/going forward?
13. What are the data management strategies?
14. How are disputes addressed?

FIGURE ES.5: Hierarchy of EM&V planning documents

TIMEFRAME		COVERAGE
Multiple Year	EM&V FRAMEWORK	Region, State, or Program Administrator
Annual or Multiple Year	PORTFOLIO CYCLE EM&V PLAN	Region, State, or Program Administrator
As Required (e.g., annual)	EVALUATION RESEARCH PLAN	Program or Portfolio
As Required (e.g., annual)	M&V PLAN	Project or Site

Executive Summary: Notes

[1] "Uniform Methods Project." (2012). U.S. Department of Energy. www1.eere.energy.gov/deployment/ump.html.

[2] Although, as discussed in this guide, this sentiment needs to be tempered with a quote that some attribute to Albert Einstein: "Not everything that can be counted counts, and not everything that counts can be counted."

Chapter 1

Introduction

1.1 GUIDE OBJECTIVE

Jurisdictions and organizations (e.g., state agencies, regulatory bodies, utilities, efficiency portfolio administrators) can use this guide as both a primer on efficiency impact evaluation and for defining their own institution-specific, general evaluation requirements as well as specific impact evaluation requirements. While each jurisdiction or entity will need to define its own evaluation requirements, this guide provides a structure, a set of evaluation approaches, suggestions on key evaluation issues, and definitions that can be applied to a variety of policy situations.

Applying the information in this guide can be particularly helpful for jurisdictions and organizations just starting or ramping up their efficiency and evaluation activities. By using standard approaches and terminology, developed through 30-plus years of efficiency program evaluation experience, costs for starting up an evaluation effort and moving "up the learning curve" can be reduced. Use of common approaches and terminology can also support comparison of efficiency programs in different jurisdictions and facilitate the implementation of "cross-border" energy efficiency and/or greenhouse gas and other air emissions mitigation programs.

1.2 SUBJECTS COVERED IN THIS GUIDE

This 2012 guide is an update to the 2007 National Action Plan for Energy Efficiency *Model Energy Efficiency Program Impact Evaluation Guide*.[3] It includes new and updated material based on feedback received on the 2007 guide and lessons learned from impact evaluations conducted during the last five years.

This guide focuses on bottom-up evaluations[4] of the impacts—primarily energy, demand, and emissions savings—of energy efficiency programs implemented in facilities, and for which energy and demand savings are the primary objectives. Therefore, the guide helps users determine the end-use fossil fuel (e.g., natural gas) and electricity savings from programs that encourage lighting, space conditioning, process approaches, and similar energy efficiency strategies in residential, commercial, institutional, and industrial facilities. In addition, while not a focus of this guide, some guidance is provided in Chapter 7 on documenting non-energy benefits and evaluating market transformation, behavior, training, and behavior-based programs. Appendix B also has sections on market effects and process evaluations, top-down evaluations, and cost-effectiveness analyses.

EXPECTATIONS AFTER READING THIS GUIDE

After reading this guide, the reader will be able to define the basic objectives, structure, and evaluation approaches that can be used to plan and conduct impact evaluations of efficiency programs. The reader will also be able to support and provide input for an energy efficiency evaluation framework (general guidance documents) and review impact evaluation plans and reports.

This guide provides the following:

- Policy-neutral[5] descriptions and guidance for planning and conducting impact evaluations of end-use efficiency programs[6] to determine energy and demand savings
- Information on determining energy and demand savings, as well as avoided emissions that result from energy efficiency programs
- Discussions about issues encountered with planning and implementing impact evaluations
- A planning process for impact evaluations including a recommended hierarchy of documents and evaluation reports
- Background on other types of energy efficiency evaluations
- A glossary of evaluation terms
- A list of other reference documents and resources on energy efficiency evaluation.

In practical terms, evaluation planners can use this guide to do the following:

- Define the questions and hypotheses that the evaluation effort is intended to address
- Identify appropriate evaluation approaches and methods that, from a budgetary perspective, balance the value of the information provided by impact evaluations with the costs to provide such information at an acceptable level of accuracy
- Set realistic expectations among the evaluation process stakeholders regarding the nature and practical value of results to be delivered, the timing of when evaluation results can be available, and the expected quality of quantitative estimates of program impacts
- Set appropriate schedules and budgets that reflect the desired level of certainty expected in the results.

PLANNING ISSUES

While reading this guide's first seven chapters, keep in mind the 14 "evaluation planning" issues listed in the Executive Summary and addressed in Chapter 8 with respect to preparing an evaluation plan.

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

- It is not sufficiently detailed to be the only resource for planning or conducting evaluations of specific programs. Rather, the guide provides high-level guidance, identifies issues, and directs 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 C does list documents where this information can be found for various program types and technologies.[7]
- It is not a guide describing how to perform feasibility studies or potential studies, which are intended to assess potential savings and benefits from future energy efficiency projects or programs, respectively. Instead, this guide can be used to help define and conduct studies that inform on what has been or is being accomplished with existing programs.

1.3 GUIDE STRUCTURE AND HOW TO USE THIS GUIDE

Table 1.1 at the end of this chapter provides a summary of the guide contents and suggestions for which chapters different audiences will find of interest.

This guide's intended audience includes the following:

- Program and evaluation managers looking for basic guidance—or a "road map"—on approaches and key issues including:
  - Defining and documenting net and gross energy and demand savings
  - Documenting avoided emissions
  - Comparing demand- and supply-side resources
- Energy system resource planners and demand forecasters looking for how end-use efficiency impact evaluation strategies and results can be effectively used in resource planning efforts
- Program designers looking to understand how their programs will be evaluated and the benefits they can receive from evaluations
- Policymakers and regulators looking for a basic understanding of evaluation objectives, processes, and issues
- Members of the energy efficiency community looking for the following:
  - Common terminology and definitions
  - A central reference that provides guidance and 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
- Expert evaluation practitioners looking to provide introductions and summaries of evaluation topics to those who do not have their expertise.

1.4 SOURCE DOCUMENTS

The information in this document is a summary of definitions, approaches, and best practices developed during more than 30 years of energy efficiency program implementation and evaluation. This experience and expertise is documented in numerous guides, protocols, papers, and reports. More information on these documents and other evaluation resources is included in footnoted references throughout the guide and in Appendix C.

USING THIS GUIDE

Policymakers and those looking for the "basics": Read the Executive Summary and first three chapters and refer to the appendices for overviews of other evaluation types, definitions, and references.

Experienced evaluation planners: Go straight to the evaluation considerations chapter (Chapter 7) and the planning chapter (Chapter 8) and use the rest of the document as a reference.

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

TABLE 1.1: Summary of Guide Content and Intended Audience for Each Part of the Guide

PART	CHAPTER	INTENDED AUDIENCE	CONTENTS
Part 1	Executive Summary	Readers interested in a brief summary and introduction to impact evaluation	The Executive Summary provides an overview of impact evaluation, with discussion of the importance and types of evaluations, the impact evaluation process, key issues, and evaluation planning.
Part 2	Chapter 1: Introduction Chapter 2: Energy Efficiency Program Evaluation Overview Chapter 3: Impact Evaluation Basics	Readers who want an overview of evaluation and the key aspects of impact evaluation	The Introduction describes the guide's objective, a review of what is and is not covered in the guide, and recommendations for how to use the guide. Chapter 2 describes the objectives of evaluation and provides definitions of different efficiency program types and evaluation categories. Chapter 3 summarizes the basics, processes, and approaches associated with impact evaluation.
Part 3	Chapter 4: Calculating Energy Savings Chapter 5: Determining Net Energy Savings Chapter 6: Calculating Avoided Air Emissions	Readers who want additional detail on impact evaluation approaches	Chapter 4 covers the three categories of approaches for determining energy and demand savings. Chapter 5 defines net savings terms and uses, briefly describes methods for determining net savings, and discusses issues associated with the use and calculation of net savings. Chapter 6 provides approaches for determining avoided air emissions associated with efficiency programs.
Part 4	Chapter 7: Impact Evaluation Considerations Chapter 8: Impact Evaluation Planning	Program implementers, evaluators, and managers/ regulators of evaluations looking for guidance on key evaluation issues and planning of evaluations as well as readers with a background in evaluation may want to go directly to these chapters	Chapter 7 provides background on topics associated with implementing impact evaluations that are not covered in other chapters, such as persistence, demand savings, controlling uncertainty, non-energy benefits, program costs, and evaluation issues associated with certain unique program types. Chapter 8 "brings it all together" and describes how the basics and details described in earlier chapters can be used to plan evaluation efforts.
Part 5	Appendix A: Glossary Appendix B: Other Evaluation Categories and Approaches Appendix C: Resources References	Readers interested in standard energy efficiency evaluation definitions and reference materials used in the evaluation industry as well as summaries of process, market evaluations, cost-effectiveness analyses, and top-down evaluation	The appendices provide resources and further background on evaluation topics.

Chapter 1: Notes

[3] 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.

[4] "Bottom-up" evaluation involves adding up the savings from measures, projects, and programs to estimate total portfolio impacts. Another evaluation category, termed "top-down," refers to approaches that rely on energy consumption data or per-unit energy consumption indicators (e.g., energy consumption per-unit of output or per person) defined for a market sector, utility service territory, or jurisdiction as the starting point for savings determination using macro-economic approaches.

[5] Because the guide is a policy-neutral document, evaluation plans must address any jurisdiction-specific policy requirements.

[6] The guide does not cover transportation-related energy efficiency programs.

[7] In addition to guidance documents, the planning and implementation of impact evaluation activities requires skilled and experienced practitioners.

Chapter 2

Energy Efficiency Program Evaluation Overview

Chapter 2 provides an overview of the three main categories of energy efficiency program evaluation. The chapter also makes the distinction between evaluations for individual energy efficiency projects and multifaceted efficiency programs. Because this guide focuses on end-use energy efficiency program evaluation, some background on different program categories is also provided. The last sections cover the importance and objectives of energy efficiency evaluation.

2.1 ENERGY EFFICIENCY PROGRAM CATEGORIES AND DEFINITIONS

Before describing the different categories of evaluation and its importance and objectives, this section provides some context regarding categories of efficiency programs that are the subject of evaluations, as well as definitions of the savings hierarchy.

2.1.1 Policy Context of Energy Efficiency Programs

The evaluation of energy efficiency programs should support the policy goals of the programs. Thus, understanding policy goals, and the context in which the programs are being implemented, affects program evaluation. Policy goals can vary widely; however, a majority of states now have policies in place that establish specific energy savings targets for energy efficiency programs provided to customers by their utilities or related organizations.[8] The following are three common ways in which states have set their efficiency goals—by legislation, regulation, or voter initiative.

- **All cost-effective energy efficiency.** This requires that the state or its utilities must acquire all energy efficiency measures that are less expensive than energy supply options. Funding for efficiency programs can be drawn from dedicated utility bill surcharges and/or the budgets that utilities would otherwise use to procure a more expensive energy supply.
- **Energy efficiency resource standard (EERS).** An EERS requires that a percentage of the resources used by utilities to supply their customers must come from energy efficiency. An EERS policy can mandate that the utility increase the percentage of energy efficiency incrementally over a number of years (e.g., 1% per year) or achieve a specific target percentage by a future date (e.g., 20% by 2025):
- **Target spending budget.** This requires that an efficiency program administrator spend a certain amount of money on energy efficiency portfolios and maximize energy or peak savings within these portfolio budgets.

2.1.2 Efficiency Program Categories

There are many types of energy efficiency programs and several approaches to differentiating them. One approach is to divide programs into two categories: voluntary and mandatory. Mandatory programs involve codes and standards that require mandated levels of efficiency in buildings and/or products (e.g., equipment or appliances). Voluntary programs involve a wide range of mechanisms to incent consumers to use energy more efficiently.

Voluntary programs can be defined as including the following subcategories and objectives:

- **Resource acquisition.** The primary objective of this program category is to directly achieve energy and/or demand savings, and possibly avoid emissions, through specific actions. This category includes activities such as rebate and direct-install programs for energy-efficient equipment, specific operational or maintenance actions (e.g., boiler tune-ups, building commissioning), and behavior-based programs that encourage consumers to adopt energy and demand savings practices. These later programs typically include outreach, education, rewards, benchmarking, and/or feedback elements.
- **Market transformation (MT).** The primary objective of this program category is to change the way in which energy efficiency markets operate (e.g., how manufacturers, distributors, retailers, consumers, and others sell and buy energy-related products and services), which tends to result in more indirect energy and demand savings. Education and training (E&T) programs and programs that support the development of or compliance with codes and standards (C&S) are examples of market transformation activities. These programs indirectly result in energy savings. To a large extent, all programs can be considered market transformation programs in that they involve a change in how energy efficiency activities take place in the marketplace.
- **Multiple objectives.** Programs can include some or all of the above-listed objectives.

SOME APPLICATIONS OF ENERGY EFFICIENCY EVALUATIONS

- Utility-administered energy efficiency programs
- Government efficiency programs, either for public 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 rely on efficiency actions
- Private company programs
- Energy service company contracts

This guide focuses on documenting the impacts of resource acquisition programs, including directly achieved energy and demand savings and related emission reductions. Section 7.10 of this guide briefly discusses evaluation of market transformation programs, including education and training and codes and standards programs. It should be noted that while a program may have just one primary objective, there are often secondary objectives that are integral to the program's overall success. This is frequently the case when resource acquisition and market transformation objectives are involved. With respect to impact evaluation, it is more important to focus on the performance goals to be assessed and establish metrics than to categorize individual program types.

End-use (consumer) energy efficiency is part of the very general category of activities known as demand-side management (DSM). Demand-side management programs are designed to encourage consumers to modify their level and pattern of energy use. Another category of DSM is demand response (DR), defined by the Federal Energy Regulatory Commission (FERC) as "a reduction in the consumption of electric energy by customers from their expected consumption in response to an increase in the price of electric energy or to incentive payments designed to induce lower consumption of electric energy."[9] Demand response programs employ energy rate design (pricing), customer incentives, and technology to enable customers to change their demand in response to system conditions or prices.

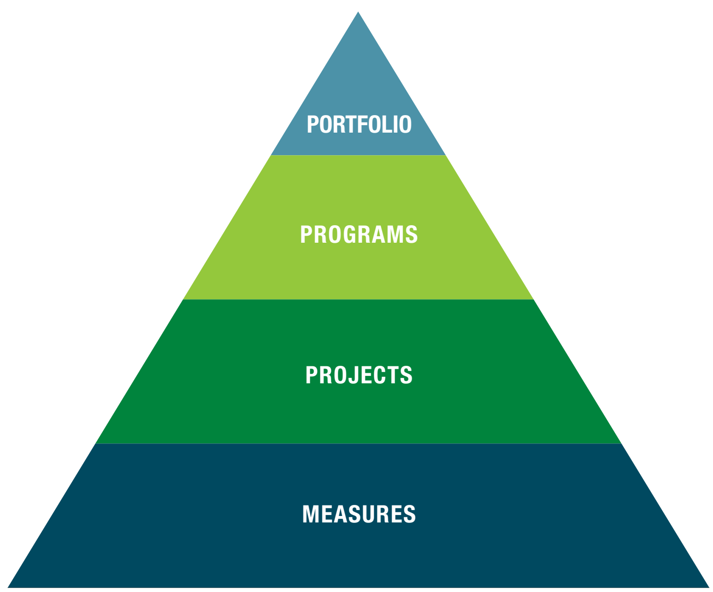
While this guide does not specifically address DR programs, the basic evaluation approaches and planning process explained here can be applied to DR with the understanding that the emphasis for DR program evaluation is demand savings. Demand savings definitions and evaluation techniques are highlighted in Sections 7.2 and 7.10.5.

2.1.3 Savings Hierarchy

The starting point for evaluating energy and demand savings, at least with bottom-up evaluation approaches, is a savings hierarchy for energy efficiency actions, as shown in Figure 2.1. This figure shows the energy efficiency actions in the following order:

- **Energy efficiency measure:** at an end-use energy consumer facility, an installed piece of equipment or system; a strategy intended to affect consumer energy use behaviors; or modification of equipment, systems, or operations that reduces the amount of energy that would otherwise have been used to deliver an equivalent or improved level of end-use service. Examples include lighting retrofits, HVAC retrofits, and commissioning.
- **Project:** an activity or course of action involving one or multiple energy efficiency measures at a single facility or site. Examples include home retrofits and commercial new construction projects.
- **Program:** an activity, strategy, or course of action undertaken by a program implementer or administrator. Each program is defined by a unique combination of program strategy, market segment, marketing approach, and energy efficiency measure(s). Programs consist of a group of projects with similar characteristics and installed in similar applications. Examples include a utility program to install energy-efficient lighting in commercial buildings, a developer's program to build a subdivision of homes that exceed common practice, or a state's effort to improve compliance with energy efficiency codes.

FIGURE 2.1: Hierarchy of energy efficiency activities



- **Portfolio:** either (1) 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 (2) the set of all programs administered by one organization, such as a utility.

2.2 PROGRAM EVALUATION CATEGORIES AND IMPACT EVALUATION DEFINITIONS

The variety of evaluation activities that are associated with energy efficiency can be categorized in several different ways, one of which is to define evaluations as either formative or outcome. Formative evaluations are associated with helping efficiency programs be as effective as possible. Outcome evaluations are associated with documenting program results. However, the most common way to categorize efficiency evaluations is as impact, process, or market evaluations. These are defined as follows (with the first two described in more detail in Appendix B).

- **Impact evaluations:** outcome evaluations of the changes attributable to an energy efficiency program. While impact evaluations usually focus on determining the quantity of changes in energy use and demand associated with a program, the calculation of non-energy benefits (or co-benefits) such as avoided emissions and job creation that directly or indirectly result from a program can also be an output of impact evaluations. Impact evaluations often support 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 (sometimes called cost-benefit) analyses may also take into account market evaluation results considering a program's short- and long-term market effects.
- **Process evaluations:** systematic assessments of an energy efficiency program. Their purpose is to document program operations and identify and recommend improvements to increase the program's efficiency or effectiveness for acquiring energy resources while maintaining high levels of participant satisfaction. For example, process evaluations can include an assessment of program delivery, from design to implementation, to identify bottlenecks, successes, failures, constraints, and potential improvements. Timeliness in identifying opportunities for improvement is essential to making corrections along the way. Process evaluations also provide a backdrop for interpreting the results of impact evaluations.

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 not only program implementation, but also 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?

Program evaluations also provide an understanding of the following:

- Program approaches that are most and least effective, as well as 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).

- **Market evaluations:** a very broad category of activities that document aspects of the marketplace with respect to energy efficiency. One particular type is a market effects evaluation, which characterizes changes in the structure or functioning of a market or the behavior of market participants that resulted from one or more program efforts. Market effects evaluations can include projections of impacts that a market could have on future energy efficiency efforts. If the evaluation's goal is to assess cost-effectiveness for stakeholders or regulators, excluding the measurement of market effects could result in underestimating (or possibly overestimating) a program's overall benefits or cost-effectiveness.

TABLE 2.1: Summary of Evaluation Categories and Types

EVALUATION CATEGORY	PHASE AT WHICH IT IS IMPLEMENTED	EVALUATION OR ANALYSIS TYPE	ASSESSMENT LEVEL
Formative	Pre-program Planning Phase	Market Assessment Analyses (includes characterization, baseline)	Market, Portfolio, Program
		Potential or Feasibility Analyses Portfolio, Program, Project	Portfolio, Program, Project
	Implementation Phase and Ongoing and/or After Program Implementation	Process Evaluations	Portfolio, Program
		Market Effects Assessments	Portfolio, Program
Outcomes	Implementation Phase—Ongoing and/or After Program Implementation	Impact Evaluations	Program, Project, Measure
		Market Effects Evaluations	Market, Portfolio
		Cost-Effectiveness Analyses	Portfolio, Program, Project

The following are example questions that could be used to determine market effects:

- Did a program encourage more vendors to offer energy-efficient products, and will there thus be future efficiency benefits associated with such increased availability of products?
- Did a voluntary incentive program prove that a new technology is viable, cost effective, and accepted by consumers, and therefore make it possible for this technology to be included in a future building code or appliance standard?

There are other types of market evaluations: market assessment studies used to determine current practices for the purposes of establishing measure, project, or program baselines; and potential studies used to estimate the technical, economic, or market-based potential of increasing the amount of energy efficiency for various products and services.

While this document focuses on impact evaluation, all types of formative and outcome evaluations are not mutually exclusive, and there are benefits to undertaking more than one type at a time and integrating the data collection and analyses functions. Thus, process evaluation and market effects evaluation often end up explicitly or implicitly bundled with impact evaluation. Table 2.1 summarizes these categories of efficiency evaluation, although not all of these different evaluations are necessary for every program or portfolio.

2.2.1 Evaluation Definitions

Evaluation is the conduct of any of a wide range of assessment studies and other activities aimed at determining the effects of a program (or a portfolio of programs). This includes understanding or documenting program performance, program or program-related markets and market operations, program-induced changes in energy efficiency markets, levels of demand or energy savings, or program cost-effectiveness. While this guide focuses on evaluations of individual programs, the basic concepts can be applied to portfolios.

Measurement and verification (M&V) is another term often used when discussing analyses of energy efficiency activities. M&V can be a stand-alone activity or it can be a subset of program impact evaluation. In either case, it is associated with the documentation of energy (and/or demand) savings at **individual sites or projects** using one or more options that can involve measurements, engineering calculations, statistical analyses, and/or computer simulation modeling. These options are defined in the International Performance Measurement and Verification Protocol (IPMVP).[10] Generally speaking, the differentiation between evaluation and project M&V is that evaluation is associated with programs (or portfolios) and M&V is associated with projects. Contractors and the owners of facilities tend to be interested in only M&V on their own project(s), while program administrators are interested in evaluation of their programs and portfolios.

As discussed in later chapters, M&V is also one of the three approaches used for program evaluation, where M&V techniques are typically used to determine the savings from a sample of projects (versus a census), with the results applied to the entire program population of projects. The other two evaluation approaches are deemed savings (which does not involve any project-specific measurement) and large-scale consumption data analysis. Both are typically applied to all of the projects (or sites) in a program.

The term **evaluation, measurement, and verification (EM&V)** is frequently seen in evaluation literature and is a catchall term for determining both program and project impacts.

2.3 OBJECTIVES AND IMPORTANCE OF EVALUATION

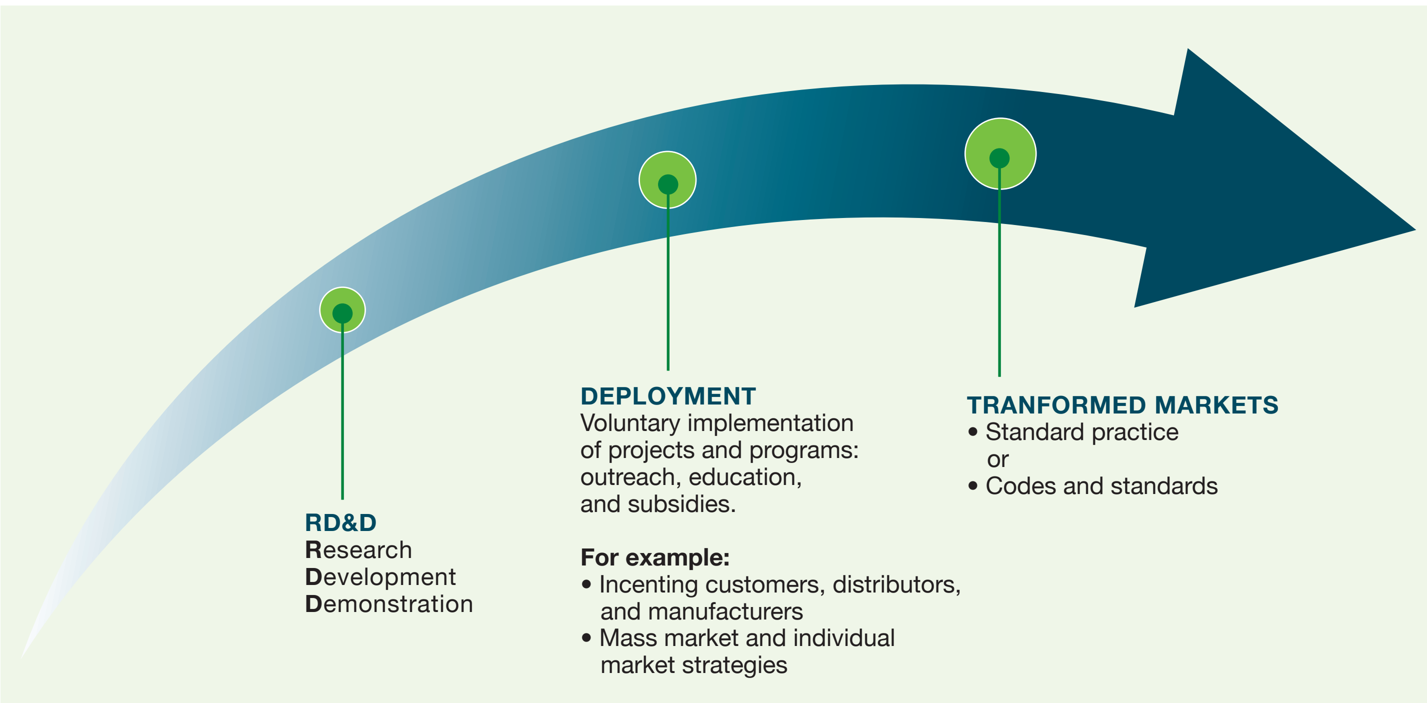
The first step in the evaluation process is to define the objectives for given evaluation activities. This leads to defining the metrics, assumptions, approaches, budgets, and other characteristics of the evaluation effort. Evaluations have three overarching objectives:

1. **Document the benefits/impacts of a program and determine whether the program (or portfolio of programs) met its goals.** Rigorous evaluations help ensure that programs are cost-effective and that benefits (e.g., energy savings, avoided emissions) are both "real" and sustained over time. This often includes assessment of compliance with regulatory requirements associated with programs funded by the public (or energy consumers). Energy efficiency impact evaluations are oriented toward developing retrospective estimates of energy savings attributable to a program, in a manner that is defensible in regulatory proceedings conducted to ensure that public funds are properly and effectively spent. Regulators support evaluation activities because of their interest in documenting total savings, assessing the cost-effectiveness of efficiency compared to generation alternatives, and assessing savings attribution (e.g., the contributions of efficiency portfolio administrators in achieving savings versus the influences of common practice, end-user self-motivation, or codes and standards). With respect to this last objective, evaluation can also be used explicitly for retrospectively determining the performance (and resulting payments, incentives, or penalties) of contractors and administrators responsible for implementing efficiency programs.
2. **Help understand why program-induced effects occurred and identify ways to improve current programs and future programs.** The role of evaluation can go well beyond simply documenting savings to actually improving current and future programs. If applied concurrently with program implementation, evaluations can provide information in real time to allow for as-needed course corrections. Evaluation fosters more-effective programs and can justify increased levels of energy efficiency investment as a long-term, reliable energy resource. Perhaps the imperative for conducting evaluation is best described by a quote attributed to John Kenneth Galbraith: "Things that are measured tend to improve."
3. **Support energy demand forecasting and resource planning by understanding the historical and future effects of energy efficiency as compared to other energy supply and demand-side resources.** As efficiency has become a more important energy resource in state and regional energy plans, an objective of impact evaluations can be to support state and regional energy forecasting and resource-planning efforts. Understanding and supporting the needs of forecasters and planners (and their data formats and definitions) can thus be an important consideration when defining end-use efficiency program evaluation metrics and reporting requirements. In addition, evaluation can support resource planning through projections of non-energy benefits, specifically emissions profiles for planning how to meet air quality and greenhouse gas mitigation objectives (see Chapter 6).

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" or is "real." Consistent, complete, accurate, and transparent evaluation mechanisms for documenting energy and demand savings, as well as non-energy benefits such as avoided emissions, address this barrier. Indeed, having effective evaluation policies, processes, and trained personnel in place to document the energy and non-energy benefits of energy efficiency programs is critical to the success of energy efficiency, emission, and climate change-mitigation programs that must prove their value and worthiness for continued investment.

Evaluation is thus not a goal unto itself; it should be viewed as one part of a continuous, and usually cyclic, process of program planning, implementation, and evaluation. The results of impact evaluation studies do not stand alone but are used as inputs into planning and improving future programs. As shown in Figure 2.2, there is a continuum of strategies associated with moving energy efficiency actions from the research, development, and demonstration (RD&D) stage to an increasing level of adoption, and to ultimately have them become standard practice or be mandated through codes and standards. With public policy orientated toward accelerating the success and fulfillment of these strategies and greater energy savings, evaluation is a tool that supports the acceleration through documentation, feedback, and energy resource planning.

FIGURE 2.2: Continuum of energy efficiency actions



Chapter 2: Notes

[8] Nowak, S.; Kushler, M.; Sciortino, M.; York, D.; Witte, P. (June 2011). *Energy Efficiency Resource Standards: State and Utility Strategies for Higher Energy Savings.* American Council for an Energy-Efficient Economy (ACEEE). Report Number U113. www.aceee.org/research-report/u113.

[9] Federal Energy Regulatory Commission (FERC). (June 2010). *National Action Plan on Demand Response.* Docket No. AD09-10. Prepared by FERC staff. www.ferc.gov/legal/staff-reports/06-17-10-demand-response.pdf.

[10] Efficiency Valuation Organization (EVO). International Performance Measurement and Verification Protocol (IPMVP). (multiple dates). www.evo-world.org. The IPMVP is an international M&V guidance document. It is discussed in Chapter 4 and references are in Appendix C.

Chapter 3

Impact Evaluation Basics

Chapter 3 is a stand-alone introduction to the basic concepts associated with energy efficiency impact evaluation. It introduces the evaluation process, the role of evaluators versus administrators of efficiency programs, and some key metrics. Also presented is the concept of savings determination based on a counterfactual situation and the fundamental resulting issue of evaluation being the balancing of evaluation costs with the reliability of savings estimates. The chapter concludes with a brief introduction to the approaches used for determining energy and demand savings: deemed savings, measurement and verification, and large-scale consumption data analysis.

3.1 IMPACT EVALUATION PROCESS

Impact evaluations determine program-specific induced effects, which include reductions in energy use (such as kilowatt-hours [kWh] and therms) and demand (kilowatts [kW]), and non-energy benefits such as avoided air emissions. The basic steps in the evaluation process are as follows:

- Set the program evaluation objectives in the context of the program policy objectives.
- Select an impact evaluation savings determination approach, define baseline scenarios, and prepare a plan that takes into account the critical issues.
- Determine energy and demand savings.
- Determine non-energy benefits (as needed).
- Report the evaluation results and, as appropriate, work with program administrators to implement recommendations for current or future program improvements and/or resource planners and demand forecasters to support their efforts.

The program evaluation process begins with defining and assessing the evaluation objectives. Well-defined objectives indicate what data need to be collected or developed during the evaluation effort and the scope and scale of effort required for meeting the objectives (e.g., the cost of obtaining the desired information, schedules, labor requirements). A key to successful evaluation is the comparison of the costs of evaluation with the value of the information that will come from the evaluation, possibly through an iterative planning process that balances cost and value.

3.1.1 Verification and Documentation

Within the impact evaluation process there tends to be two types of major activities: verification and documentation. These may be more aptly called "verifying the potential to generate savings" and "documenting (determining) the actual savings." To illustrate the difference between the two, consider a project involving replacement of 100-watt incandescent lamps with 23-watt compact fluorescent lamps (CFLs).

EVALUATION PLANNING ISSUES

Chapter 8 discusses the evaluation planning process. The planning process is used to decide which (or which combination) of the evaluation metrics and approaches defined in this guide should be used.

Verification would involve confirming that the replaced lamps are 100 watts, that the new CFLs are 23 watts, and that the CFLs were installed and are working. As a result of this verification, it can be confirmed that the project has the potential to save energy, with the amount of energy saved dependent on how many hours the lamps operate. Determining how many hours the lamps operate, in this case, would be the "documentation" activity (or what some might call the "measurement," although as discussed in Section 3.3, savings cannot be literally measured).

In this guide, *verification* is defined formally as an independent assessment that a program has been implemented per the program design. For example, the objectives of measure installation verification are to confirm (1) the installation rate (number of units installed), (2) that the installation meets reasonable quality standards, and (3) that the measures meet the program eligibility requirements and are operating correctly with the potential to generate the predicted savings. For some programs, it may be that verifying the potential to generate savings is all that is needed to meet the evaluation objectives, while in many other situations, both verification and documentation of the actual savings value will be required.

Verification may include one-time or multiple activities over the estimated life of the measures and can overlap with commissioning (Cx) or retro-commissioning (RCx) activities, which have similar objectives—to ensure the installed equipment is working correctly and per design. However, verification may not go "as far" as Cx or RCx to ensure the operation of the energy efficiency measure(s).

BASIC IMPACT EVALUATION CONCEPTS

- Impact evaluations are used for determining achieved program effects.
- 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; in other words, they appropriately balance risk management, uncertainty, and cost considerations.

3.1.2 The Evaluator's Role and Reporting of Savings

There are several entities involved in the actual implementation of efficiency projects and programs, and they may each conduct their own impact evaluation activities. These entities include the end-use energy consumers who have projects installed in their facilities, designers, contractors, and program implementers and administrators. Some of these entities may only be interested in verification activities or just their own individual projects' measurement and verification, while others will conduct complete program evaluations. The following definitions help to explain these roles.

Entities:

- **Administrator:** an entity selected by a regulatory or other government organization to contract for and administer an energy efficiency portfolio within a specific geographic region and/or market. Typical administrators are investor-owned or public utilities, nonprofits, and state government agencies. An administrator could also be a private entity that hires a company, such as an energy services company (ESCO), to implement its efficiency program(s).
- **Implementer:** an entity selected and contracted with, or qualified by, a program administrator to provide products and/or services to consumers, either directly or indirectly.
- **Independent third-party evaluator:** an entity that conducts evaluations and is designated to be independent of the implementer and administrator.

Reported Savings:

- **Projected savings:** values reported by an implementer or administrator prior to the time the subject energy efficiency activities are completed. These are typically estimates of savings prepared for program and/or portfolio planning purposes.
- **Claimed savings:** values reported by an implementer or administrator after the subject energy efficiency activities have been completed.
- **Evaluated savings:** savings estimates reported by an independent third-party evaluator after the subject energy efficiency activities and an impact evaluation have been completed. These can differ from claimed savings in that an independent third-party evaluator, to an agreed-to level of rigor, has conducted evaluation and/or verification activities.

The implementers and/or administrators usually prepare their projected savings estimates and claimed savings estimates. They will thus conduct their own evaluation activities, using their own evaluation staff or consultants, for purposes such as confirming any incentive payments to program participants or contractors and preparing documentation for internal and external reporting.[11] If an independent third-party evaluator is used, that evaluator will then conduct some level of evaluation (verification only or verification and their own data collection/analysis to determine savings) for preparation of its own evaluation reports and a realization rate[12] comparing evaluated savings with projected savings estimates and/or claimed savings estimates. Of course, the evaluator can and should be brought into the process before any of this work is conducted to participate in defining the roles and responsibilities of the administrators, implementers, and evaluators as well as reporting requirements. The designation of "independent" and "third-party" is determined by those entities involved in the use of the evaluations and may include

QUALITY ASSURANCE GUIDELINES

The impact evaluation approaches described in this guide are based on new and unique analyses of energy and demand savings. However, often there is documentation on energy and demand savings from analyses prepared independently of the subject impact evaluation. Even though such documentation was not necessarily prepared per predetermined evaluation requirements, it may be sufficient for meeting the evaluation objectives. Using existing documentation in combination with quality assurance guidelines (QAG) can significantly reduce overall program/evaluation costs. Essentially, a QAG can help determine whether indicated savings, and the assumptions and rigor used to prepare the existing documentation, can be used in place of new evaluation efforts.

evaluators retained, for example, by the administrator or a regulator. Defining the relative roles of the administrator, implementer, and independent third-party evaluator is another important activity of the planning process. Section 8.3.2 of this guide discusses evaluator roles and selection.

3.2 ENERGY AND NON-ENERGY BENEFIT EVALUATION METRICS

For energy and demand savings (and conceptually for non-energy benefits) the primary metrics are known as gross energy savings and net energy savings. In this guide, based on industry standard practice, gross and net energy (and demand) savings are defined as follows:

- **Gross energy 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. This is the physical change in energy use after taking into account factors not caused by the efficiency actions (e.g., changes in weather or building occupancy).
- **Net energy savings:** the change in energy consumption and/or demand that is attributable to a particular energy efficiency program. Estimating net energy savings typically involves assessing free ridership and spillover, although this guide discusses additional considerations. In the efficiency industry, *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 (i.e., the participant would have undertaken the energy-saving activity anyway). *Spillover* refers to the program-induced adoption of measures by nonparticipants and participants who did not claim financial or technical assistance for additional installations of measures supported by the program. For instance, a participant undertakes additional energy efficiency measures due to positive experience with the program, or a nonparticipant undertakes such measures based on observing a program participant's results. Net savings estimates also sometimes include consideration of market effects.

The difference between these two metrics is associated with (1) attribution of the savings—in other words, the determination of whether the savings were caused by the program being studied (entirely or partially) or by other influences such as prior year programs or other programs/influences operating at the same time as the program; and (2) differences in how different entities (e.g., regulatory bodies) define net and gross savings. Approaches for determining gross and net savings are summarized in this chapter, with additional information provided in Chapters 4 and 5.

SAVINGS

Savings, or more accurately stated "savings estimates," from energy efficiency measures, projects, programs, and portfolios are reported at various times in the lifecycle of the efficiency activity and with varying degrees of certainty. Savings are most commonly reported at two major milestones—prior to and after the implementation of the activity. Savings can also be indicated as first-year, annual, and/or lifetime energy or demand savings values. They also can be indicated as gross savings and/or net savings values. Different jurisdictions currently have different names for savings reports, what they contain, and whether and what adjustments or evaluation activities take place between preimplementation and postimplementation. Ideally, the terms and methods in this guide should be applied. However, whenever savings are reported, it is critical that the basis for the values indicated be made clear.

When energy or demand savings are reported, they are typically estimated for the first year of a program, for a specific number of years (e.g., 5, 10, 15), or for the life of the program's measures. *Measure life* is the length of time that an energy efficiency measure is expected to be functional and generating savings. It is a function of equipment life and measure persistence. *Equipment life* is the number of years that a measure is installed and will operate until failure. *Measure persistence* refers to the duration of an energy-consuming measure, taking into account business turnover, early retirement of installed equipment, and other reasons measures might be removed or discontinued. Measure life is sometimes referred to as *expected useful life (EUL)*.

The other two main metric categories are non-energy benefits and cost-effectiveness.

- **Non-energy benefits (NEBs):** the identifiable—although sometimes unquantified—non-energy impacts associated with program implementation or participation; also referred to as *non-energy impacts* (NEI) or co-benefits. Examples of NEBs include environmental benefits, productivity improvements, jobs created, reduced program administrator debt and disconnects, and higher comfort and convenience levels of participants. The value of NEBs is most often positive, but may also be negative (e.g., the cost of additional maintenance associated with a sophisticated energy-efficient control system), which is why some practitioners prefer the term NEIs. Potential benefits of efficiency to the energy system (e.g., price stability, grid reliability, and power quality) are "energy-related" but are also often put into this general category of NEBs. The primary NEB addressed in this guide is avoided air emissions (see Chapter 6). Section 7.9 has more information on calculating non-energy benefits.

- **Cost-effectiveness:** an indicator of the relative performance or economic attractiveness of any energy efficiency investment or practice relative to energy supply resources. It is another metric that is commonly used when reporting the results of impact evaluations. In the energy efficiency field, the present value of the estimated benefits produced by an energy efficiency program is compared with the estimated total costs of the program in order to determine whether 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 or from a program participant perspective).

3.3 FUNDAMENTAL ENERGY EFFICIENCY IMPACT EVALUATION CONCEPTS: THE COUNTERFACTUAL AND MANAGING UNCERTAINTY

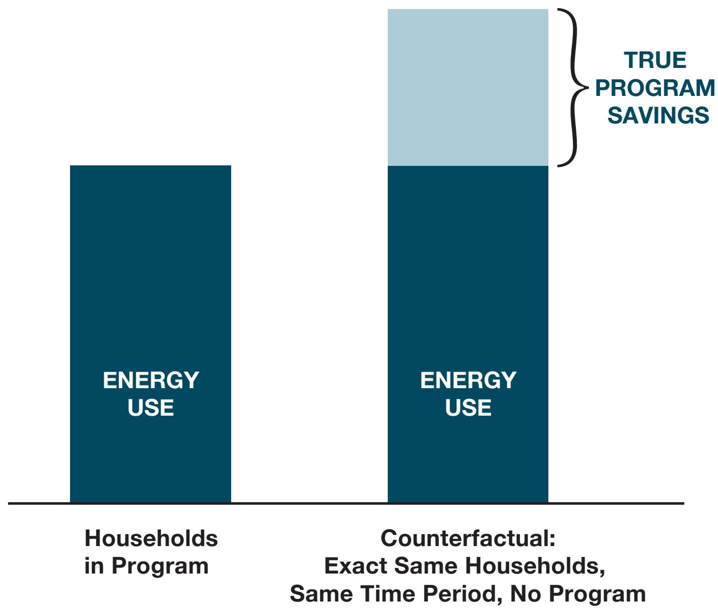
3.3.1 The Counterfactual

In theory, the true energy savings from an energy efficiency program is the difference between the amount of energy that participants in the program use relative to the amount of energy those same participants would have used had they not been in the program (during the same time period). This baseline is called the counterfactual scenario (see Figure 3.1). However, in practice, we can never observe how much energy those participants would have used had they not been in the program, because at any given time a participant must either be in the program or not. Thus, there is no direct way of measuring energy (demand) savings, because (1) it is not possible to measure a participant's energy use with and without the program, at the same time; and (2) one cannot measure the absence of energy use.

Defining this counterfactual scenario represents the fundamental concept and the greatest challenge to documenting the benefits of energy efficiency. This challenge is met with impact evaluations measuring energy consumption—but, the savings themselves will always be estimates. The savings estimate is the difference between (1) actual energy consumption after a project or program is implemented, and (2) what energy consumption would have occurred during the same period, by the same participants, had the efficiency project/program not been implemented.

The graph in Figure 3.2 summarizes this estimation process. The blue line represents energy use of a building before, during, and after an efficiency project is implemented. This energy use can be known

FIGURE 3.1: True program savings: the counterfactual for a residential household energy efficiency program



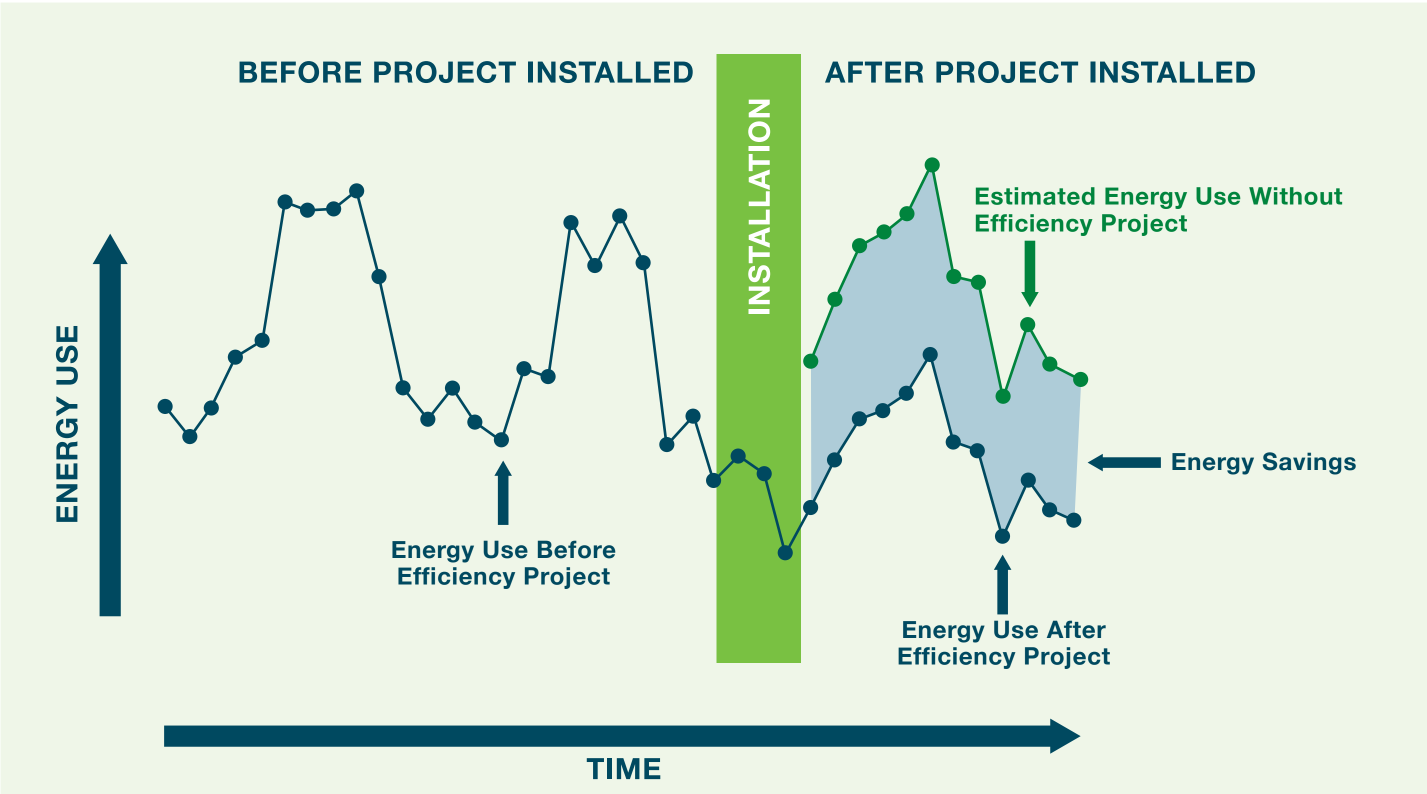
(e.g., through measurement), but to determine the savings (the blue shaded area), the energy use that would have occurred without the project (the green line) has to be estimated in order to determine a value for energy savings.

As discussed in Chapter 2, an objective of program evaluation is to produce energy and demand savings values (and, as desired, associated non-energy benefits). However, as noted above, these values are always going to be estimates; the use of these 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. Most of the remainder of this guide describes the approaches, issues, and planning processes that should be considered when addressing the counterfactual challenge and undertaking impact evaluations.

3.3.2 Baselines

The baseline is the counterfactual scenario, determined on the basis of a number of considerations: the evaluation approach being used, the type of project being implemented, site-specific issues, and broader policy-orientated considerations. These considerations usually result in one of three different types of baselines being selected for the impact evaluation: existing conditions, common practice, or

FIGURE 3.2: Energy consumption before, during, and after a project is implemented



codes/standards. Baselines are discussed throughout this guide with respect to the evaluation approaches and planning for evaluation activities. The first section of Chapter 7 provides more information on selecting baselines.

3.4 DETERMINING ENERGY AND DEMAND SAVINGS

The three impact evaluation approaches used to determine energy and demand savings can be grouped into two conceptual frameworks: noncontrol group approaches and control group approaches.

- **Noncontrol group approaches.** These are the deemed savings and M&V approaches defined below and in greater detail in Chapter 4. With these approaches, pre-project (or pre-program) baseline energy use is defined using one or more of a variety of different methods. This baseline is compared with post-project (or post-program) energy use measurements or assumptions to estimate savings. These noncontrol group approaches generate estimates of gross savings, which require adjustments to determine net savings.
- **Control group approaches.** These are the large-scale consumption data analysis approaches that are also described in Chapter 4. With these approaches, a comparison group's energy use is compared with the energy use of program participants. These approaches, in most cases, generate estimates of net savings, taking into consideration free ridership and participant spillover, but do not take into account nonparticipant spillover and long-term market effects, which some jurisdictions include in the net savings determination.

3.4.1 Noncontrol Group Impact Evaluation Approaches

The following are brief summaries of the two noncontrol group approaches: measurement and verification and deemed savings.

- **Measurement and verification (M&V).** Measurement and verification is the process of using measurements to reliably determine energy and/or demand savings created within an individual facility. The International Performance Measurement and Verification Protocol (IPMVP), an international M&V guidance document, defines four M&V options used in the efficiency industry: two end-use metering approaches, energy use data (billing data) regression analysis, and calibrated computer simulation.

- **Deemed savings.** Deemed savings are based on stipulated values, which come from historical savings values of typical projects. A typical source of such historical values are prior year M&V or large-scale consumption data analysis studies. Unlike the M&V approach, with the use of deemed savings, there are no (or very limited) measurement activities; instead, only the number of measures implemented is verified (e.g., number of motors installed correctly, number of point-of-sale CFLs that were sold). This approach is only valid for projects with fixed operating conditions and well-known, documented stipulation values. This approach involves multiplying the number of installed measures by the estimated (or deemed) savings per measure.

A variant of deemed savings is the deemed savings calculation, which is one or more agreed-to (stipulated) engineering algorithm(s) used to calculate the energy and/or demand savings associated with an installed energy efficiency measure. These calculations may include stipulated assumptions for one or more parameters in the algorithm, but typically they require users to input data associated with the actual installed measure into the algorithm(s).

MEASUREMENT AND VERIFICATION VERSUS DEEMED SAVINGS

For simple, well-defined, 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 period for which savings are reported (e.g., first year or lifetime), deemed savings can help alleviate some of the guesswork in program planning and design; in effect, they minimize one type of risk by providing certainty. However, deemed savings can result in another form of risk if not properly developed and applied—overestimates or underestimates of savings if the projects or products do not perform as expected. This can occur, for example, if the deemed savings value is incorrectly calculated or the deemed savings value was simply applied to the wrong type of application.

Measurement-based approaches are more appropriate for more complex efficiency projects or for project with significant savings variability (i.e., those with a significant amount of savings, or “risky” savings, or with no history or analysis or metering on which to base a deemed savings value). Measurement-based approaches are also more rigorous than deemed savings approaches and involve site data collection during the period of evaluation for at least the most critical variables. 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 in Chapter 4.

3.4.2 Control Group Approaches: Large-Scale Consumption Data Analysis (Randomized Controlled Trials and Quasi-Experimental Methods)

A reliable (precise, unbiased) approach for estimating energy savings from efficiency programs is to measure the difference between the energy use of facilities (e.g., houses) participating in a program (the “treatment group”) and that of a similar comparison group of non-participating facilities (the “control group”) during the same period of time. The two generic categories of control group approaches are randomized controlled trials and quasi-experimental methods:

- **Randomized controlled trials (RCTs).** In an RCT, a study population (e.g., single-family houses in Seattle that have electric heat) is defined and randomly assigned to either the treatment group or the control group. Energy use (consumption) data must be collected for all of the project sites in the treatment and control group in order to estimate energy savings. The energy savings estimate is then calculated by comparing the difference between the measured energy use (or preferably the difference between the measured change in energy use) of the treatment households and the energy use of the control households during the same period.
- **Quasi-experimental methods.** Unlike RCTs, with quasi-experimental methods the assignment of the control group is not totally random. Thus, quasi-experimental methods, relative to RCTs, often suffer from selection bias and may produce biased estimates of energy savings. However, because of the difficulty and costs of conducting RCTs, quasi-experimental approaches are more common than RCTs, with perhaps the most common being the “pre-post” approach. With this approach, sites in the treatment group after they were enrolled in the program are compared with the same sites’ historical energy use prior to program enrollment. In effect, this means that each site in the treatment group is its own nonrandom control group.

With these approaches, statistical analyses are conducted on the energy use data (typically collected from the meter data reported on monthly utility bills) and other important independent variable data (e.g., weather) for those in the control and treatment groups. These approaches are primarily used for programs with relatively homogenous participants and measures, when project-specific analyses are not required or practical, but could, at least in theory, be considered for every type of program. Example applications are large-scale weatherization programs and residential behavior-based programs.

3.4.3 Approaches for Determining Net Energy and Demand Savings

For the noncontrol group approaches where gross savings are determined, the difference between net and gross savings is specified as a net-to-gross (NTG) ratio. The following common approaches are used to determine net savings:

- **Stipulated net-to-gross ratios.** These are ratios that are multiplied by the gross savings to obtain an estimate of net savings and are based on historical studies of similar programs. Sources of stipulations can cover a wide range, from simply using "negotiated guesses" to historical values to structured expert judgment panels. This is the least expensive approach.
- **Self-reporting surveys and enhanced self-reporting surveys.** Enhanced surveys include interviews and documentation review and analysis. These are moderately expensive approaches.
- **Panel of trade allies.** A significant number of trade allies provide information on their recent projects, whether the projects are in the subject program or not, to assess the program's impact on incented and nonincented program energy efficiency measures.
- **Large-scale consumption data analysis approaches** (randomized controlled trial methods and quasi-experimental methods). When a control group of nonparticipants is used, the savings indicated are "net" of free riders and participant spillover. These are discussed in Chapter 4.
- **Cross-sectional studies.** These studies are comparisons of market share of targeted technologies or behaviors between a baseline area not served by the program and the area served by the program.
- **Top-down evaluations.** These evaluations use state, regional, or national data at a sector level to assess the extent to which markets for energy-efficient products and services have been affected by programs.

It is not unusual for combinations of these approaches to be used. For example, rigorous randomized controlled trials may be used every three years, with self-reported or deemed NTG ratios used for the other program years. More information about determining net savings is provided in Chapter 5.

THE COUNTERFACTUAL SCENARIO

A counterfactual analysis occurs when a person modifies a factual antecedent (a thing or event that existed before or logically precedes another) and then assesses the consequences of that modification. A person may imagine how an outcome could have turned out differently if the factual situation, or what led to it, did not occur. This may seem daunting, but for energy efficiency impact evaluations, this is simply defining what the energy use (or demand, emissions, number of jobs, etc.) would have been if the program had not been implemented.

The fact that energy and demand savings, as well as related non-energy benefits, from efficiency efforts cannot be directly measured results in analyses based on a counterfactual scenario. It is counterfactual because savings are not measured, but rather estimated to varying degrees of accuracy by comparing energy consumption after a program is implemented (the reporting period) with what is assumed to have been the energy consumption (and demand) in the absence of the project program (the baseline or the counterfactual scenario). The baseline and reporting period energy use and demand are compared and adjusted so that only program effects are considered when determining savings. These adjustments are a major part of the evaluation process and can vary from one program type to another and from one evaluation approach to another.

Chapter 3: Notes

[11] In many administrator organizations, these internal evaluations are called "measurement and verification" or just "verification" and are conducted by the same team that implements the programs.

[12] Realization rate is used in several contexts for comparing one savings estimate with another. The primary and most meaningful application is the ratio of evaluated gross savings to claimed gross savings (versus comparing net and gross savings estimates, which is best defined with a net-to-gross ratio). Basis for the ratio not being 1.0 can include several considerations such as the following: (1) adjustments for data errors, (2) differences in implemented measure counts as a result of verification activities, and/or (3) other differences revealed through the evaluation process, such as with respect to baseline assumptions.

Chapter 4

Calculating Energy Savings

As discussed in Chapter 3, there is no direct way of measuring energy savings, because one cannot measure the absence of energy use. However, the absence of energy use (i.e., savings) can be estimated. Within the efficiency evaluation industry there are three generic classifications of savings determination approaches, and the following sections describe each of them in more detail than what was summarized in earlier chapters: measurement and verification, deemed savings, and large-scale consumption data analysis (with the use of control groups). These descriptions are intended to be overviews, with additional information sources referenced in Appendix C. The last section of this chapter describes some criteria for selecting an approach. Supporting information in Chapter 5 provides information on determining net savings. Chapter 7 provides information on other impact evaluation topics related to calculating energy savings (and demand and non-energy benefits), and Chapter 8 discusses planning evaluation efforts.

4.1 MEASUREMENT AND VERIFICATION APPROACH

Measurement and verification (M&V) is the determination of gross energy savings at individual sites or projects using one or more methods that can involve measurements in combination with engineering calculations, statistical analyses, and/or computer simulation modeling.

Measurement and verification is a project-based approach to determining savings. Its genesis was in the efficiency performance contracting industry, starting in the 1970s. With performance contracting, a contractor implements an efficiency project for a client via a contractual arrangement that includes a savings (performance) guarantee or an arrangement that payment to the contractor is dependent on the savings achieved. These arrangements required "measurement and verification" to determine what level of savings were being achieved at the client's facility. By the early 1990s, the growth of the performance contracting industry was constrained, in part, by the lack of robust methodologies for verifying project savings, which, in turn, restricted the ability of the project finance industry to participate in the market. To this end, the efficiency industry (represented by the National Association of Energy Service Companies)[14] worked with the U.S. Department of Energy (DOE), the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), and other stakeholders in the mid-90s to develop guidelines, which became the basis for the International Performance Measurement and Verification Protocol (IPMVP),[15] described below.

As publicly funded efficiency programs became more prevalent in the 1990s, the energy efficiency industry adopted M&V techniques and concepts into the growing number of evaluations that were being conducted for efficiency programs. Thus, M&V is now used extensively as a program evaluation approach, primarily for "custom projects" whose savings are dependent on the technologies applied and/or the situations in which they are applied. For program evaluation, M&V involves the following activities:

- Selecting a representative sample of projects in a specific efficiency program, although in some programs all of the projects may be selected for M&V (a census)
- Determining the savings of each project in the sample, which is done with M&V activities consisting of the following:
- Development of a M&V plan
  - Meter calibration and installation and, for long-term measurements, maintenance
  - Data gathering and screening
  - Computations of savings with measured data
  - Quality assurance reviews and reporting
- Applying the sample projects' savings to the entire population (i.e., the program).

UNIFORM METHODS PROJECT

A source of more detailed impact evaluation information is DOE's Uniform Methods Project (UMP),[13] which provides model evaluation plans for specific energy efficiency measures and project categories. These UMP documents contain additional information and specific examples that apply the concepts presented in this guide and include examples of the three impact evaluation approaches presented in the following sections.

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). As part of a broader verification process, field inspections 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 project 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.

M&V also includes 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.

The industry's primary M&V resource is the IPMVP, which is an international end-use, energy efficiency M&V guidance document. It provides a framework for conducting M&V, and most important, defines four M&V options that are used in the efficiency industry. The options involve metering of all relevant parameters, metering of key parameters, energy use (billing) data regression analysis, and/or computer simulation. Complementing the IPMVP are the Federal Energy Management Program (FEMP) *M&V Guidelines*.[16] These guidelines, and companion technical notes, provide more details on M&V for specific measure and technology applications. A third important M&V resource is ASHRAE Guideline 14 (2002): *Measurement of Demand and Energy Savings*.[17] This guideline provides technical detail on subjects such as metering. A new version of Guideline 14 is expected to be available in early 2013.

The following subsections provide material mostly from the 2010 version of the IPMVP (EVO 10000—1:2010). They cover the basic M&V concept (algorithm) for calculating gross energy savings at the project level and introduce the four IPMVP Options—A, B, C, and D—using the descriptions found in the IPMVP.

With M&V, energy (and/or demand) savings are determined for a project by comparing energy use (and/or demand) before and after implementation of the energy efficiency measures. Thus, the following fundamental algorithm applies for energy (and demand) savings:

Energy savings = (Baseline energy use) – (Reporting period energy use) ± (Baseline adjustments)

- **Baseline energy use:** the energy consumption that would have occurred without implementation of the energy efficiency activity. When discussed in terms of specific projects, where energy saving is the metric of interest, it is sometimes called *preinstallation energy use.*
- **Reporting period energy use:** the energy consumption that occurs within the time frame following implementation of an energy efficiency activity during which savings are to be determined. When discussed in terms of specific projects, it is sometimes called *postinstallation energy use*.
- **Baseline adjustments:** factors that modify baseline energy or demand values to account for independent variables (such as weather) that influence energy use. These adjustments account for conditions in the reporting period that are different from the conditions during the baseline period but are not a result of the project activity. They distinguish properly determined savings from a simple comparison of energy use before and after implementation of a program. By accounting for independent variables that are, or are not, 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 include the following:
  - Weather corrections (e.g., if the program involves heating or air-conditioning systems in buildings)
  - Occupancy levels and hours (e.g., if the program involves lighting retrofits in hotels or office buildings)
  - Production levels (e.g., if the program involves energy efficiency improvements in factories).

One of the other considerations for adjustments is unique to program evaluation. In almost all bilateral performance contracts between a contractor and its client, the baseline is considered to be whatever "exists in the facility" before the efficiency measures are implemented. However, in many programs using public or energy consumer funds, the baseline may be defined as common practice or as required by a code or standard (in order to not give credit for what would have been done without the project activity). In this situation, the savings calculated from an end-use consumer perspective may be different from the savings determined for a government agency or regulatory body. Thus, a single project may involve two different sets of savings calculations—one that supports the contract between the energy services contractor and the customer (using a baseline of existing conditions in the customer facility), and a second that supports the customer's claim to a utility rebate (using a baseline of current codes and standards). See Chapter 7 for more discussion of the nuances of selecting baselines.

4.1.1 Measurement and Verification Approaches

If M&V is a part of the evaluation process, at least some M&V details will need to be specified in the evaluation planning documents, and if sampling is used, a basis for selecting the sample of specific project sites at which M&V activities will take place will be needed. In addition, as M&V is a project-specific approach to evaluation, 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 and generic method plans.

- **Prescriptive method plans.** For project types 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 standardized algorithms. The FEMP *M&V Guidelines* contain prescriptive approaches for documenting savings for several common energy efficiency measures, as does ASHRAE Guideline 14. The DOE UMP materials also include several standardized approaches to documenting savings. The UMP documents do so in the context of M&V as an evaluation approach and are thus more germane to the subject of this guide than the FEMP and ASHRAE documents which are more specific.
- **Generic method plans.** There are conceptual approaches applicable to a variety of project types 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 M&V approaches.

One of the other important 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., 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 method developed to estimate them under the savings determination process.

The four IPMVP options (A, B, C, and D) provide a flexible set of methods for evaluating project energy/demand savings. Having four options provides a range of methods for determining energy/demand savings with varying levels of savings certainty and cost. A particular option is chosen based on the specific features of each project, including the following:

- Energy efficiency measure technologies employed and the end uses in which they are applied
- Complexity, particularly in terms of interactive effects with multiple measures and energy-using systems
- Potential for changes in key factors during the baseline and/or reporting periods
- Uncertainty of the project savings as compared to the value of project savings
- Value of understanding the performance of the measures (e.g., for a new technology).

The options differ in their approach to the level, duration, and type of baseline and reporting period measurements. The options also differ in terms of measurement boundaries:

- Measurement boundaries with Options A and B are made at the end use, system level (e.g., lighting, HVAC).
- Measurement boundaries with Options C and D are at the whole-building or whole-facility level.

Additionally, they differ in terms of type of measurements and their duration:

- Option A involves using a combination of both stipulations and measurements of the key factors needed to calculate savings in engineering models. Data collection tends to involve either spot-measurements or short-term measurements.[18]
- Options B and C involve using spot, short-term, or continuous measurements 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.

The four generic M&V options are summarized in Table 4.1.

TABLE 4.1: IPMVP Options Summary

IPMVP OPTION	HOW SAVINGS ARE CALCULATED	TYPICAL APPLICATIONS
A. Retrofit Isolation: Key Parameter Measurement Savings are determined by field measurement of the key performance parameter(s), which define the energy use of the energy conservation measures (ECMs) affected system(s) and/or the success of the project. Measurement frequency ranges from short-term to continuous, depending on the expected variations in the measured parameter and the length of the reporting period. Parameters not selected for field measuring are estimated. Estimates can be based on historical data, manufacturer's specification, 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 calculation of baseline and reporting period energy from: • Short-term or continuous measurement of key operating parameter(s); estimated values • Routine and nonroutine adjustments as required.	A lighting retrofit in which power draw is the key performance parameter that is measured periodically. Estimate operating hours of the lights based on building schedules and occupant behavior.
B. Retrofit Isolation: All Parameter Measurement Savings are determined by field measurement of the energy use of the ECM-affected system. Measurement frequency ranges from short-term to continuous, depending on the expected variations in the savings and length of the reporting period.	Short-term or continuous measurement of baseline and reporting period energy, and/or engineering computations using measurements of proxies of energy use. Routine and nonroutine adjustments as required.	Application of a variable-speed drive and controls to a motor to adjust pump flow. Measure electric power with a kW meter installed on the electrical supply to the motor, which reads the power every minute. In the 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.
C. Whole Facility Savings are determined by measuring energy use at the whole-facility or subfacility 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 a simple comparison or regression analysis. Nonroutine adjustments as required.	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 a subfacility. Simulation routines are demonstrated to adequately model actual energy performance measured in the facility. This option usually requires considerable skill in calibrated simulation.	Energy use simulation, calibrated with hourly or monthly utility billing data. (Energy end-use metering may be used to help refine input data.)	Multifaceted energy management program affecting many systems in a facility but where no meter existed in the 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: Efficiency Valuation Organization (EVO). International Performance Measurement and Verification Protocol. (2010). IPMVP, EVO 10000—1:2010. www.evo-world.org.

4.1.2 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 at the "end-use" (e.g., ventilation system, lighting system) level. With this option, key performance parameters or operational parameters can be 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 the development of estimates of energy and demand savings based on the following:

- 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 fixtures 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 calculate energy use for the subject end use (e.g., cooling system or lighting system). 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 facility historical data or manufacturer's data on the affected baseline and/or project equipment. It is appropriate to use a stipulated factor only if supporting data demonstrate that its value is not subject to fluctuation over the term of analysis and it is demonstrably applicable to the project.

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 and the variability of its savings (e.g., Option A is not a very reliable option for energy management system retrofits that involve complex building, environment, user, and operator interactions). Thus, the savings from a simple lighting retrofit (less complex) may be more accurately determined with Option A than could 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 because the cost of measuring one or two parameters is usually less than measuring all of the parameters. However, because some stipulation is allowed under this option, care is needed to review the engineering design and installation to ensure that the stipulations are realistic, applicable, and achievable (i.e., the equipment can truly perform as assumed). This can be done through "desk reviews" of data but is more reliably done as part of the verification process, where site inspections check the efficiency measure characteristics and collect site data. At defined intervals during the reporting period, the installation can be reinspected to verify the equipment's continued existence and its proper operation and maintenance. Such reinspections will ensure continuation of the potential to generate predicted savings and validate stipulations and prior savings estimates.

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

Option B, as with Option A, involves project or system-level (end-use) M&V assessments with performance and operational parameters measured at the component or system level. Option B also involves procedures for verification activities 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 any major factors that would have a significant influence on energy or demand savings.

Thus, Option B requires additional and often longer-term measurements compared to Option A. These include measurements of both equipment operating characteristics (as may be required under Option A) and relevant performance factors (which may not be required under Option A). Commonly measured parameters include operating hours for lighting and HVAC equipment, wattage for lighting and HVAC equipment, and flow rates and pressure for various compressed-air applications. 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 determining energy savings using Option B can be more difficult than doing so with Option A. And, as noted above for Option A, the difficulty and cost increase as measurement complexity and savings variability increase. The savings, however, are typically more reliable than those determined with Option A.

4.1.4 M&V Option C: Whole-Facility Analyses

Option C involves use of whole-building meters or submeters to assess the energy performance of a building or facility. These meters are typically the ones used for utility billing, although other meters, if properly calibrated, can also be used. With this option, energy consumption from the baseline period is compared with energy consumption data (usually derived from energy bills) from the reporting period. Option C also involves procedures for verification activities that are the same as Option A.

Whole-building or facility-level metered data are evaluated using techniques ranging from simple bill data 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 substantive explanatory (independent) variables that affect energy consumption need to be monitored during the performance period. Substantive variables may include weather, occupancy schedules, industrial throughput, control set points, and operating schedules. Most applications of Option C require at least 9 to 12 months of monthly baseline (preinstallation) meter data and at least 9 to 12 months of monthly data from the reporting period (postinstallation).

All end-use technologies can be verified with Option C. However, this option is intended for projects in which savings are expected to be large enough to be distinguishable 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.

Option C is the most common form of M&V for some multi-measure building energy efficiency retrofits in the performance contracting industry (although ESCOs use Option A as much as possible and then use Options B, C, or D as needed). For programs targeting integrated whole-building approaches to energy efficiency, utility bill analysis can be used to statistically evaluate persistence. One useful tool for this purpose is the U.S. Environmental Protection Agency's (EPA's) ENERGY STAR® Portfolio Manager (see sidebar on the next page).

It should be noted that the term *billing analysis* is often used generically to describe any analytic methodology used to determine project or program energy savings based on the use of the energy consumption data contained in consumer billing data. It compares billing data from program participants over a period of time before the energy efficiency measures are installed at customer sites to billing data for a comparable period of time afterward. If used to describe a project-based measurement and verification approach, it is equivalent to the IPMVP Option C: Whole Facility Analysis. If billing analysis is used to describe a program-based evaluation approach, it is comparable to the large-scale consumption data analysis approach, which involves billing data from both participants and nonparticipants (control group).

4.1.5 M&V Option D: Calibrated Simulation

Option D involves calibrated computer simulation models of systems, system components, or whole-facility (usually residential or commercial buildings) energy consumption to determine project energy savings. While in theory simulation can involve the use of any computer analysis tools, such as spreadsheets, these calibrated simulations are typically associated with complex building analysis tools that model heating, cooling, lighting, ventilation, and other energy flows as well as water use and sometimes onsite air emissions. Examples of such programs are DOE-2 and EnergyPlus.[19] The quality of the savings estimate depends on how well the simulation models are calibrated and how well they reflect actual performance.

INTERACTIVE FACTORS AND LEAKAGE

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 effects beyond a single building and start to affect 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.

U.S. EPA'S PORTFOLIO MANAGER

One tool that can be used to analyze facility utility billing meter data is U.S. EPA's Portfolio Manager (PM). Over 300,000 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. It 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 consumption. www.energystar.gov/portfoliomanager.

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, although simulation results (reporting period) to simulation results (baseline) are also used to determine savings.

Models are often calibrated by comparing simulation results with historical data to ensure that the models have accurately captured the operating characteristics of the building. 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 preinstallation 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.

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 in which there is a high degree of interaction among installed energy systems or 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.

4.2 DEEMED SAVINGS APPROACH

Deemed savings values, also called *stipulated savings values*, are estimates of energy or demand savings for a single unit of an installed energy efficiency measure that (1) has been developed from data sources (such as prior metering studies) and analytical methods that are widely considered acceptable for the measure and purpose, and (2) is applicable to the situation being evaluated. Individual parameters or calculation methods can also be deemed; for example, effective useful life of a measure or the annual operating hours of light fixtures in an elementary school classroom. Common sources of deemed savings values are previous evaluations and studies that involved actual measurements and analyses. Deemed savings values are used for both planning and evaluation purposes.

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. Many performance contracting projects document their savings with deemed savings, and it is also a popular evaluation approach for many efficiency programs because of both the relatively low cost of using deemed savings and the certainty of savings values that all parties can rely on for their own purposes.

The use of deemed values in a savings calculation is thus essentially an agreement between the involved 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 and performance, satisfactory commissioning results, and sufficient equipment or system maintenance),

BUILDING ENERGY SIMULATION PROGRAMS

For about 40 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, visit http://simulationresearch.lbl.gov/resources.

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. Savings can also be verified for "persistence" with periodic inspections that verify that the retrofits are still in place and functioning.

A variant of deemed savings is the *deemed savings calculation*, which is an agreed-to (stipulated) set of engineering algorithm(s) used to calculate the energy and/or demand savings associated with an installed energy efficiency measure. These calculations are developed from common practice that is widely considered acceptable for the subject measure and its specific application. It may include stipulated assumptions for one or more parameters in the algorithm, but typically it requires users to input data associated with the actual installed measure into the algorithm(s).

The use of deemed savings is quite popular for evaluations of energy consumer-funded programs. The following is from a recent survey of evaluation practices in the United States:

> We found that nearly all states (36 states, 86%) use some type of deemed values in the evaluation framework. In terms of what types of values are "deemed," we found 35 states (97% of those responding to this question) deem savings amounts for particular measures, 32 states (89%) deem the "lifetime" over which to claim savings for particular measures, and 20 states (65%) deem free-ridership or net-to-gross factors.
>
> We also inquired about the source of the deemed values used by the states. It appears that there is a lot of "borrowing" going on within the industry. Twenty-six states (70%) cite the use of sources or databases from other states. In nine states, the utilities develop and file certain key deemed values, and in two states, the Commission is responsible for developing the deemed values. In most states (28 states, 80%), the results of their own in-state evaluations are used to modify and update deemed values over time.[20]

Deemed savings values and deemed savings calculations are usually documented in a database in formats from spreadsheets to online searchable databases. A term of art for such databases is *technical reference manuals (TRMs)*. These are resource documents that include energy efficiency measure information used in program planning and energy efficiency program reporting. It can include savings values for measures, measure life information, hourly load shapes of savings, engineering algorithms to calculate savings, impact factors to be applied to calculated savings (e.g., net-to-gross values), source documentation, specified assumptions, and other relevant material to support the calculation of measure and program savings as well as the application of such values and algorithms in appropriate applications. For example, a value for operating hours in an elementary classroom with no summer hours should not be applied to a high school classroom with summer sessions.

As of the date of this guide's publication, there are approximately 17 TRMs in use across the United States. These include state and regional TRMs, which are listed in Appendix C. In a recent SEE Action report on TRMs,[21] it was shown that these resources are very valuable, but there are a wide variation in methodologies for estimating savings and the actual values. Some TRMs include information based on prior year evaluations including, in some cases, rigorous metering and analysis, and thus these TRMs contain robust (reliable) savings values. Many others have values based on analyses (e.g., using computer simulations or engineering algorithms), with consideration in their calculations of waste heat factors, in-service rates, and partial load factors. The transparency and level of detail regarding methods and assumptions also ranges from substantial to minimal.

Thus, as would be expected when using any assumptions or stipulated values in an analysis, caution should be used to understand the sources of such values and ensure that the assumptions that went into determining a value are applicable to the situation (e.g., measures, measure delivery mechanism, facility types) being evaluated. Deemed values, if used, should be based on reliable, traceable, and documented sources of information, such as the following:

- 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.

In addition, it is good practice to have an ongoing process in place to assess the validity of deemed savings values, such as an annual or biennial process to update TRMs. In particular, check to see if the assumptions used to determine deemed savings values are valid in the years after their initial determination (e.g., the applicability and validity of assumed code requirements or standard practices).

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 to estimate savings, the result would be underestimated savings if lamps were only installed in high-use areas (kitchens) and overestimated savings if lamps were only installed in low-use areas (closets). Thus, the value of deemed savings depends not only on the validity of the value used, but on whether the value is applied correctly (i.e., it must be based on the use conditions as well as the technology).

In summary, 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 be very useful for program planning purposes and can reduce M&V costs, create certainty, and simplify evaluation procedures. Improperly used, they can give evaluation results an inappropriate aura of authority. Deciding whether parameters could be stipulated requires that users understand how they will affect savings, judge their effect on the uncertainty of results, and balance the costs, risks, and goals of the program being evaluated.

4.3 LARGE-SCALE CONSUMPTION DATA ANALYSIS (RANDOMIZED CONTROLLED TRIALS AND QUASI-EXPERIMENTAL METHODS)

As indicated in Chapter 3, a reliable approach for estimating energy savings from efficiency programs is to measure the difference between the energy use of those facilities (e.g., houses) participating in a program (the "treatment group") and the energy use of a comparison group of nonparticipating facilities (the "control group") that are similar to those in the participant group during the same period of time. The difference between the energy use of the treatment group and the control group facilities can be attributed to three sources:

1. The true impact of the program
2. Preexisting differences between households in the treatment and control group, which is called "bias" or "selection bias"
3. Inherent randomness.[22]

There are two generic types of large-scale consumption data analyses discussed in this guide: randomized controlled trials (RCTs) and quasi-experimental methods (QEMs). Both involve the use of control groups, but RCTs provide less biased and typically more precise results than QEMs, although they may require more effort to implement. For efficiency programs, RCTs and QEMs are particularly useful for programs in which there are a relatively large number of participants with similar characteristics (e.g., low-income, single-family houses in a particular city). Currently, these methods are primarily used for evaluations of residential behavior-based programs and whole-house retrofits and weatherization programs.

4.3.1 Randomized Controlled Trial (RCT) Methods

In an RCT, first a study population is defined (such as homes eligible for a residential weatherization program or a behavior-based program in a particular city). Then the study population is randomly assigned to either the treatment group or control group.[23] Energy use data must be collected for all facilities in both the treatment group and control group in order to estimate energy savings. Measured energy use typically comes from utility meter or billing data, often in hourly or monthly increments.

The energy savings estimate is then calculated in one of two ways (options): (1) by comparing the difference between the measured energy use of the treatment group households[24] and the energy use of the control households during the evaluation study period, or (2) by comparing the energy use reduction (i.e., the change in use from before the study period to after the study period) between the treatment households and the control households. Thus, savings from the program are essentially the difference in changes in energy use (positive or negative) between the treatment group and the control group. Between the two options listed above, the second is preferable so that the differences between the control group and treatment group with respect to their differences in energy use before and after the program implementation (i.e., differences in differences) are determined.

Random assignment is a defining aspect of RCTs. It means that each facility (e.g., household) in the potential program participant population is randomly assigned to either the control or treatment group (i.e., those that will be in the program) based on a random probability, as opposed to being assigned to one group or the other based on some characteristic of the facility or participant (e.g., location, energy use, or willingness to sign up for the program). Randomization thus eliminates preexisting differences that are observable (e.g., energy use or household floor area) as well as differences that are typically unobservable (e.g., attitudes regarding energy conservation, number of occupants, expected future energy use, and occupant age) unless surveyed. Thus, because of this random assignment, an RCT control group is an ideal comparison group: it is statistically identical to the treatment group in that there are no pre-existing differences between the two groups, which means that selection bias is eliminated. Randomized controlled trials can also be used for various program enrollment options, including opt-in, opt-out, and a randomized encouragement design that does not restrict program participation.[25]

It is worth pointing out two net savings aspects that RCTs address: free riders and participant spillover. This is one of the main benefits of an RCT over traditional evaluation methods. Thus, if net savings are defined for a program evaluation as the gross savings plus consideration of free riders and participant spillover, RCTs deliver estimates of net savings.

RCTs address the free-rider concern because the treatment and control groups each contain the same number of free riders through the process of random assignment to the treatment group or the control groups. When the two groups are compared, the energy savings from the free riders in the control group cancel out the energy savings from the free riders in the treatment group, and the resulting estimate of program energy savings is an unbiased estimate of the savings caused by the program (the true program savings). Participant spillover is also automatically captured by an RCT design.

An RCT design also addresses rebound effects or take-back during the study period, which can occur if consumers increase energy use (e.g., as a result of a new device's improved efficiency). Rebound is sometimes, although not often, a consideration in determining savings persistence. Rebound effects after the study period can be accounted for with an RCT if the data collection is continued for the time under consideration.

However, free riders and participant spillover are not the only factors differentiating gross and net savings. The following are some other net-to-gross considerations that are not addressed by either RCTs or the QEMs about to be discussed:

- Nonparticipant spillover issues, in which a program influences the energy use of non-program participants (although there are some specialized techniques that can determine nonparticipant spillover)[26]
- Natural gas-related changes in energy use if only electricity consumption is measured
- Long-term market effects such as changes in efficiency product pricing and availability (unless control and treatment groups are maintained over long periods) or the influence of a program on establishing an energy efficiency measure as common practice or part of a code or standard.

4.3.2 Quasi-Experimental Methods (QEMs)

Other evaluation design methods that use nonrandomized control groups are called quasi-experimental methods. With these methods, the control group is not randomly assigned. Thus, quasi-experimental methods often suffer from selection bias that may produce biased estimates of energy savings—sometimes very biased results that result in unreliable savings estimates. However, in some specific cases in which RCTs are not feasible, quasi-experimental approaches can still provide reliable results (especially compared with deemed savings values and certain M&V approaches).

The sections that follow provide brief discussions of some common QEMs.

Pre-Post Energy Use Method

Probably the most common quasi-experimental method is to compare the energy use of participants in the treatment group after they were enrolled in the program to the same participants' historical energy use prior to program enrollment. In effect, this means that each participant in the treatment group is its own nonrandom control group. This is called a *pre-post, within subjects*, or *interrupted time series design analysis*.

The challenge in using this method is that there are many other factors (independent variables) that may influence energy use before, during, and after the program that are not captured with this method, resulting in biased savings estimates. Some of these factors, such as differences in weather or number of occupants, can be measured and reliably accounted for in the analysis. However, other factors are less easily observed and/or accounted for. For example, the economy could worsen, leading households and businesses to decrease energy use (even if there were no program); prices for energy can change, affecting energy use; and cultural norms could change—perhaps, say a pop culture icon could suddenly decide to advocate for energy efficiency.[27]

To minimize bias when using the pre-post calculation method, it is necessary to include a regression analysis that discerns and controls for the impact of other influences (e.g., economic recession) that may affect energy use over time compared with the impact of the efficiency program. Simple comparison before and after energy use is not acceptable.

Matched Control Group Method

If it is not possible to create a randomized control group, then savings estimates could be calculated by constructing a nonrandom control group made up of participants that are as similar to the treatment group as possible. The challenge with a matched control group method is that participants and their facilities have both observable characteristics (e.g., level of energy use, ZIP code, presence of central air conditioning) that could potentially be matched, as well as unobservable characteristics (e.g., energy attitudes or propensity to opt in to an energy efficiency program) that are harder or impossible to match.

Variation in Adoption

The *variation in adoption* approach takes advantage of variation in the timing of program adoption in order to compare the energy use of participating facilities that optin to the program with the energy use of facilities that have not yet opted in but will ultimately opt in at a later point. It relies on the assumption that in any given month, participants that have already opted in and participants that will opt in soon share the same observable and nonobservable characteristics. Thus, this method creates a control group that is very similar to the treatment group over time in both observable and unobservable characteristics, and therefore is likely to result in less bias than matched control or pre-post methods.

Regression Discontinuity Method

Among the quasi-experimental methods, regression discontinuity typically yields the most unbiased estimate of energy savings. However, it is also the most complicated method, as it requires knowledge of econometric models and often requires field conditions that allow the evaluator to use this analytic technique. Therefore, it is neither common nor always practical. This method works if the eligibility requirement for households to participate in a program is a cutoff value of a characteristic that varies within the population. For example, households at or above a cutoff energy consumption value of 900 kWh per month might be eligible to participate in a behavior-based efficiency program, while those below 900 kWh are ineligible. In this case, the households that are just below 900 kWh per month are probably very similar to those that are just above 900 kWh per month. Thus, the idea is to use a group of households right below the usage cutoff level as the control group and compare changes in their energy use to households right above the usage cutoff level as the treatment group. This method assumes that the program impact is constant over all ranges of the eligibility requirement variable that are used in the estimation (e.g., that the impact is the same for households at all levels of energy use). In addition, regression discontinuity relies on the eligibility requirement being strictly enforced.

4.3.3 Analysis Techniques

All of the methods described above (RCTs and QEMs) use one of a number of different analysis techniques, including regressions and other statistical and econometric methods, to analyze measured energy use data and to control for variations in independent variables.

With regression analyses, an equation or group of equations that model the relationship between the dependent variable and one or more important independent variables is defined. Dependent variables are those that are modeled to be influenced by the independent variables. Independent variables are the variables that are not influenced by other variables in the model, and are assumed to affect or determine the dependent variables, and are thus the inputs to an analysis. Independent variables include both the variable(s) of interest (e.g., a variable that indicates which customers experience a critical peak pricing event during which hours) as well as control variables that seek to observe and account for other factors that may influence the dependent variable (e.g., the average temperature). Which independent variables are relevant to calculating energy savings? Often, this is decided by common sense, experience, program characteristics, or budget considerations (with respect to how many variables can be measured and tracked), but it also can be determined through field experiments and statistical tests. For weather data, the most common independent control variable, there is a wide range of public and private data sources.

In the case of energy efficiency analyses, the output of a regression analysis is a coefficient that estimates the effect of independent variables (e.g., a program) on the dependent variable (energy or demand consumption and/or savings). The analysis itself is done with a computer model, which can be anything from a spreadsheet tool to sophisticated proprietary statistical modeling software.

4.4 SELECTING AN ENERGY 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 generic impact approaches discussed in the proceeding sections are applicable. More information on planning evaluation activities is included in Chapter 8.

One criterion that is applicable across all of the approaches is evaluator experience and expertise. Thus, a common requirement for selecting an approach is that the evaluator has expertise with the approach selected. A related requirement is that the resources required for that approach—such as the data, time, and budget—are available.

4.4.1 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, such as when there are no deemed savings values available that are applicable to the given measure or combination of measures, or when per-project results (savings) 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. Table 4.3 indicates factors that affect the cost of implementing each option. Table 4.4 and Figure 4.1, both from the 2010 version of the IPMVP, indicate key project characteristics that are better for the different options and a flowchart summarizing the selection of M&V options, respectively. These tables and the figure are included as the last pages of this chapter.

4.4.2 Deemed Savings Approach

The *deemed savings* approach is most commonly used for programs that involve simple new construction, or for retrofit energy efficiency measures with well-defined applications and savings calculations that have been verified with data. Examples might be a residential second-refrigerator recycling program or a CFL giveaway for residential utility customers. In each of these two examples, an assumption would be made about the baseline and energy savings, as well as perhaps the life of the measures (e.g., 15 years for a refrigerator and 10,000 hours for a CFL). The deemed savings values would have to be defined for specific applications (e.g., second refrigerators in single-family housing) and program delivery mechanisms (point of sale or direct install). The deemed savings values might also have a “vintage” in that they are valid for certain years of a program or for certain vintage baseline equipment.

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 that have been verified with data, and are not subject to significant variation in savings due to changes in independent variables. The stipulated values do not significantly increase the uncertainty of the evaluation metrics.

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, such as information prepared by implementers and/or administrators. 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 (QAGs) can save significant costs for the program sponsor—and perhaps encourage participation in the program if a portion of evaluation costs is 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.

- The uncertainty associated with savings estimates is low and/or the risk of under- (or over-) estimating savings is low. That is, the project’s likelihood of success is high.
- Documented, reliable, and applicable 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 ensure they are properly installed and have the potential to generate savings (rather than having rigorously determined energy savings).

Assessing a few key aspects of the project can help in making decisions about whether to use deemed savings or deemed calculations. 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. The “rules of thumb” are as follows:

- 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 should 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.

4.4.3 Large-Scale Consumption Data Analysis Approach

These approaches are used for programs that have many participants that share many common characteristics, such as single-family detached homes in a particular community with residents of similar economic demographics. These can be equipment retrofits, new construction, or behavior-based programs. Because of the requirement for a large amount of data, this approach is almost always used with residential programs, such as a weatherization program with thousands of homes being retrofitted with a variety of measures (e.g., insulation, weather stripping, low-flow showerheads, and CFLs).

In general, the large-scale consumption 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 consumers or facilities that participate in the program are known).
- The program has a relatively large number of participants (i.e., probably more than 100).
- At least one year's worth of baseline and reporting period energy consumption data are available for both the treatment group and the control group. If an RCT method is used, a shorter data period may be adequate.
- If an RCT method is not used, then there are observable similarities between participants, or relatively homogenous subgroups of participants can be formed with similar facility and energy efficiency measure characteristics.
- Either the program design is such that the target population for the program can be randomly divided into participants and nonparticipants (for randomized controlled trial methods), or sufficient data about the characteristics of the participants are available for selecting an appropriate control group for quasi-experimental methods. This is a particularly important criterion with respect to control group members, as they typically do not have an incentive to provide data to the evaluator.
- Independent, third-party evaluators select the treatment and control group, rather than program implementers, to minimize the potential for "gaming."

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 adjust-ments are anticipated • Baseline measurement data do not exist or are prohibitively expensive to collect

Source: Efficiency Valuation Organization (EVO). (2010). International Performance Measurement and Verification Protocol. IPMVP, EVO 10000—1:2010. www.evo-world.org.

TABLE 4.3: Factors Affecting the Cost of Each M&V Option

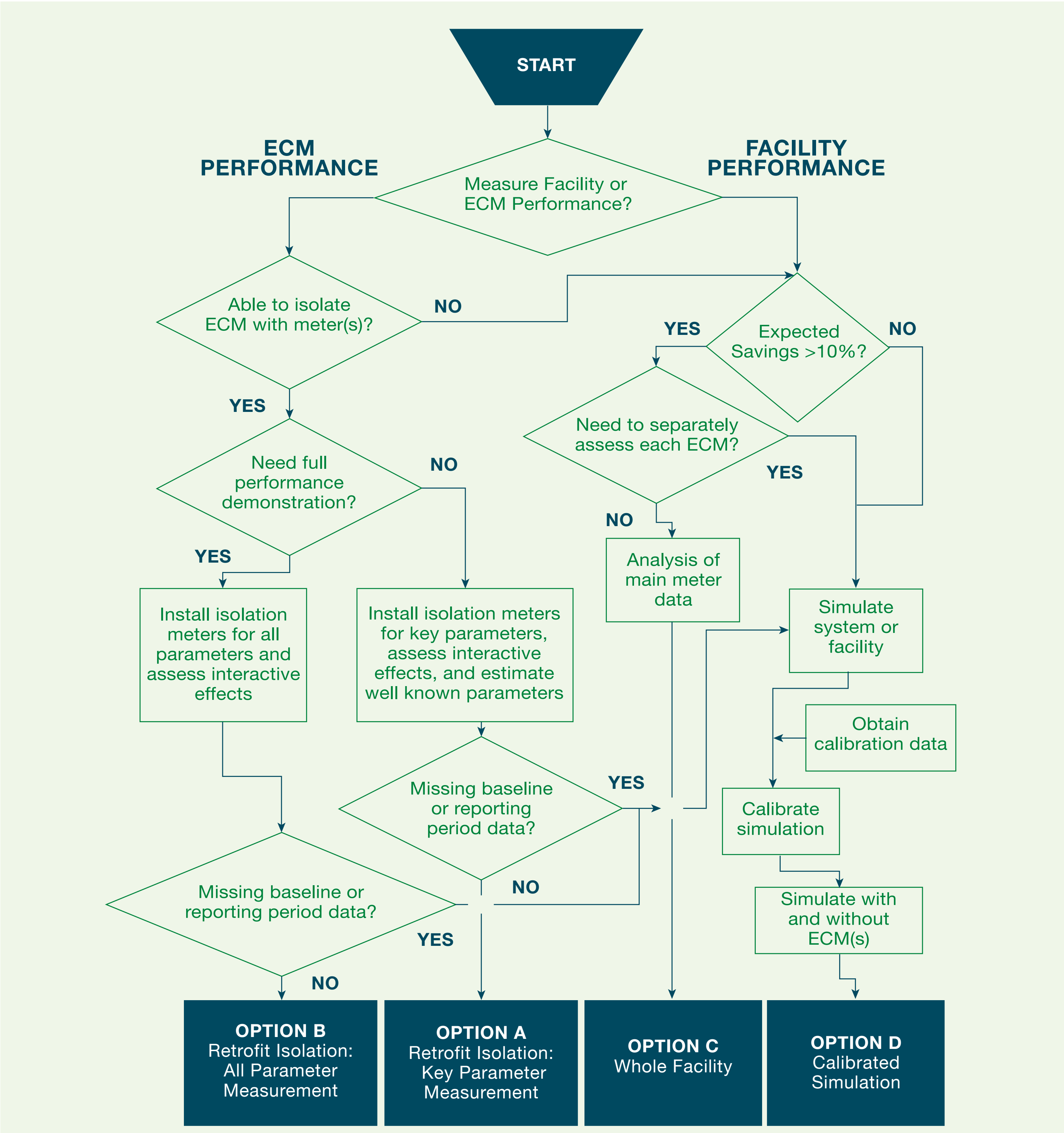
OPTION A Retrofit Isolation – Key Parameter Measurement	OPTION B Retrofit Isolation – All Parameters Measurement	OPTION C Whole Facility	OPTION D Calibrated Simulation
• Number of measurement points • Complexity of deriving the stipulation • Frequency of postretrofit inspections	• Number of points and independent variables measured • Complexity of measure-ment system • Length of time measure-ment system maintained • Frequency of postretrofit inspections	• Number of meters to be installed and/or analyzed • Number of independent variables used in models	• Number of meters to be installed and/or analyzed • Number of independent variables used in models • Effort required for calibration of models

TABLE 4.4: Key Project Characteristics and Applicability for Different M&V Options

ECM PROJECT CHARACTERISTIC	SUGGESTED OPTION			
	A	B	C	D
Need to assess ECMs individually	X	X		X
Need to assess only total facility performance			X	X
Expected savings less than 10% of utility meter	X	X		X
Multiple ECMs	X		X	X
Significance of some energy driving variable is unclear		X	X	X
Interactive effects of ECM are significant or unmeasurable			X	X
Many future changes expected with measurement boundary	X			X
Long-term performance assessment needed	X		X	
Baseline data not available				X
Nontechnical persons must understand reports	X	X	X	
Metering skill available	X	X		
Computer simulation skill available				X
Experience reading utility bills and performing regression analysis available			X	

Source: Efficiency Valuation Organization (EVO). (2010). International Performance Measurement and Verification Protocol. IPMVP, EVO 10000—1:2010. www.evo-world.org. In this table, ECM stands for energy conservation measure, which is equivalent to energy efficiency measure as used in this guide.

FIGURE 4.1: M&V option selection flowchart



Source: Efficiency Valuation Organization (EVO). (2010). International Performance Measurement and Verification Protocol. IPMVP, EVO 10000—1:2010. www.evo-world.org. In this table, ECM stands for energy conservation measure, which is equivalent to energy efficiency measure as used in this guide.

Chapter 4: Notes

[13] "Uniform Methods Project." (2012). U.S. Department of Energy. www.eere.energy.gov/deployment/ump.html.

[14] NAESCO. www.naesco.org.

[15] Efficiency Valuation Organization (EVO). International Performance Measurement and Verification Protocol (IPMVP). (multiple dates). www.evo-world.org. The Efficiency Valuation Organization (EVO) is a non-profit organization that also offers a Certified M&V Professional Certificate Program.

[16] Federal Energy Management Program (FEMP). (April 2008). *M&V Guidelines: Measurement and Verification for Federal Energy Projects*. www1.eere.energy.gov/femp/pdfs/mv_guidelines.pdf. Other related M&V documents can be found at http://mnv.lbl.gov/keyMnVDocs/femp.

[17] American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). (2002). Guideline 14: *Measurement of Demand and Energy Savings*. www.ashrae.org. Also ASHRAE. (2010). *Performance Measurement Protocols for Commercial Buildings*. www.ashrae.org/standards-research--technology/special--project-activities. These documents provide a standardized, consistent set of protocols for facilitating the comparison of measured energy, water, and indoor quality performance of commercial buildings.

[18] *Spot measurements* are one-time measurements, such as 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. Utility billing meters provide continuous measurements and are the typical measurements used with IPMVP Option C. The use of longer-term measurements can help identify under-performing energy efficiency projects, which in turn can lead to improvements in their performance.

[19] DOE-2 information can be found at http://apps1.eere.energy.gov/buildings/tools_directory/alpha_list.cfm and EnergyPlus information can be found at http://apps1.eere.energy.gov/buildings/energyplus/energyplus_about.cfm. An extensive list of building energy simulation programs can be found at: http://apps1.eere.energy.gov buildings/tools_directory/alpha_list.cfm.

[20] Kushler, M.; Nowak, S.; Witte, P. (February 2012). *A National Survey of State Policies and Practices for the Evaluation of Ratepayer-Funded Energy Efficiency Programs*. American Council for an Energy-Efficient Economy (ACEEE). Report Number U122. www.aceee.org/research-report/u122.

[21] State and Local Energy Efficiency Action Network. (June 2011). *Scoping Study to Evaluate Feasibility of National Databases for EM&V Documents and Measure Savings*. Prepared by Jayaweera, T.; Haeri, H.; Lee, A.; Bergen, S.; Kan, C.; Velonis, A.; Gurin, C.; Visser, M.; Grant, A.; Buckman, A.; The Cadmus Group Inc. www.eere.energy.gov/seeaction/pdfs/emvscoping__databasefeasibility.pdf.

[22] Randomness can be a factor in these analyses. For example, it could be that for the chosen population, during the time interval the energy use was monitored, the households in the treatment group randomly happened to lower their energy use at around the same time that the program started. The precision of an estimate of energy savings quantifies the effect of this inherent randomness and allows us to decide whether it is a problem or not. Bias and precision, as well as the general subject of certainty of savings estimates, is discussed in Section 7.4.

[23] The control and treatment groups could contain equal sizes of households, or the control group could be bigger or smaller than the treatment group. It is only necessary to keep a control group that is sufficiently large to yield statistical significance of the savings estimate.

[24] Because households are the most common example, the rest of this section uses households as the example unit of analysis.

[25] State and Local Energy Efficiency Action Network. (May 2012). *Evaluation, Measurement, and Verification (EM&V) of Residential Behavior-Based Energy Efficiency Programs: Issues and Recommendations*. Prepared by Todd, A.; Stuart, E.; Schiller, S.; Goldman, C.; Lawrence Berkeley National Laboratory. www1.eere.energy.gov/seeaction/pdfs/emv_behaviorbased_eeprograms.pdf. A complete source of information on use of RCTs (and QEMs) for evaluation of energy efficiency programs, particularly behavior-based residential programs. Much of the text in this section is from that report.

[26] It is possible to explicitly conduct an experiment to determine the spillover effects. For example, an experiment could observe the impacts of the intervention for households in the experiment as well as others and compare the impacts in two or more communities. This type of experiment is often used in medicine and epidemiology.

[27] However, in some programs, such as critical peak pricing or critical peak rebates, a pre-post method may be less biased. This is because with such programs the experimental factor (e.g., the call to curtail electricity use or a price signal) can be presented repeatedly so that the difference between the energy use when it is present and when it is not present is observable.

Chapter 5

Determining Net Energy Savings

This section describes an important, and perhaps the most controversial of, impact evaluation topics—the determination and use of net savings. Net (energy and/or demand) savings are the changes in energy consumption or demand that are attributable to an energy efficiency program. The net-to-gross (NTG) ratio is the portion (it can be less than or greater than 1.0) of gross savings (those that occur irrespective of whether they are caused by the program or not) that are attributed to the program being evaluated.

Determining net savings involves separating out the impacts that are a result of influences other than the program being evaluated, such as consumer self-motivation or effects of other programs. The controversy of net savings can be summarized as follows:

Given the range of influences on consumers' energy consumption—and the complexity in separating out both short-term and long-term market effects caused by the subject programs (and other programs)—attributing changes to one cause (i.e., a particular program) or another can be quite complex.

This controversy is compounded by a lack of consensus by energy efficiency policymakers and regulators as to which short-term and long-term market influences and effects should be considered when determining net savings and the role of net savings in program design, implementation, and "crediting" of savings to program administrators.

The following subsections start with classic energy efficiency industry definitions of net savings, including the commonly used factors that differentiate net from gross savings, and a discussion of the uses of net savings determinations. Next is a description of the approaches used to determine net savings. The final subsection discusses some of the issues and nuances of net savings.

5.1 DEFINITIONS AND USES OF NET SAVINGS

5.1.1 Defining Net Savings

The energy efficiency community agrees on the basic definition of *net savings*: the total change in energy use (and/or demand) that is attributable to an energy efficiency program. However, as noted in a recent regional scoping paper on net savings, the operational definition of net savings—essentially what factors are considered when determining net savings—is not unanimously agreed to, with different factors being applied in different programs and jurisdictions.[28] Factors that cause the difference between net and gross savings, implicitly or explicitly, include free ridership, participant and non-participant spillover, and induced market effects. These factors may be considered in how a baseline is defined (e.g., common practice) and/or in adjustments to gross savings values.

To help understand these terms, the following are some of the related definitions:

- **Free rider:** a program participant who would have implemented the program's measure(s) or practice(s) in the absence of the program. Free riders can be (1) total, in which the participant's activity would have completely replicated the program's intended actions; (2) partial, in which the participant's activity would have partially replicated the program's actions; or (3) deferred, in which the participant's activity would have partially or completely replicated the program's actions, but at a future time beyond the program's time frame.
- **Spillover (participant and non-participant):** the reduction in energy consumption and/or demand caused by the presence of an energy efficiency program, beyond the program-related gross savings of the participants and without financial or technical assistance from the program. There can be participant and/or non-participant spillover. *Participant spillover* is the additional energy savings that occur when a program participant independently installs incremental energy efficiency measures or applies energy-saving practices after having participated in the efficiency program as a result of the program's influence. *Non-participant spillover* refers to energy savings that occur when a program non-participant installs energy efficiency measures or applies energy savings practices as a result of a program's influence. Sometimes the term *free drivers* is used for those who have spillover effects.
- **Market effect:** a change in the structure of a market or the behavior of participants in a market that is reflective of an increase (or decrease) in the adoption of energy-efficient products, services, or practices and is causally related to market intervention(s) (e.g., programs). Examples of market effects include increased levels of awareness of energy-efficient technologies among customers and suppliers, increased availability of efficient technologies through retail channels, reduced prices for efficient models, build-out of efficient model lines, and—the end goal—increased market share for efficient goods, services, and design practices.

In terms of how different jurisdictions define net savings, and which of the above factors are included, a 2012 American Council for an Energy-Efficient Economy study indicated the following:[29]

> We asked states what they used when they report their energy savings results, and found that 21 states (50%) said they reported net savings, 12 states (29%) said gross savings, and 9 states (21%) said they report both (or use one or the other for different purposes). We explored the net savings issue in a little more detail, and asked whether states made specific adjustments for free riders and spillover. Interestingly, while 28 states (67%) indicated they make an adjustment for free riders, only 17 states (44%) make an adjustment for free drivers/spillover.

It is important to recognize that the study survey did not specify any particular definition of what qualifies as net or gross savings. Rather, they allowed states to categorize their own approach.

From a summary paper of evaluation practices, the traditional evaluation approach to net savings assumes a measurement boundary around a program, as shown in Figure 5.1.[30] The participant savings, which are the documented savings from the efficiency measures installed through the program's transactions with the customer, are shown in the shaded circle. Any of those savings attributable to free riders are subtracted out (inner circle), leaving the shaded donut of net participant savings. If there are spillover savings, either from additional measures installed by participants without program incentives or from measures installed by non-participants who were indirectly influenced by participants' actions, these may be added to the total. Finally, as the programs begin to affect the market (e.g., by inducing retailers to sell only efficient equipment in response to market demand), there may be additional savings (i.e., the market effects). These additional savings would expand the shaded area of countable program savings.

However, as also pointed out in the above-referenced paper, with the evolution of integrated portfolios of programs targeted at broad and deep savings goals, the situation looks more like Figure 5.2. Note that the overlaps may, in practice, be even greater than this simplified diagram would suggest. Often, the participant savings overlap between program offerings, and one program's participants may be another program's free riders. The overlaps can also occur over time, with one program's influence extending to the next generation of programs. The point of the diagram is to suggest that the simple evaluation strategy of drawing a boundary around each program, over a single program cycle, will encounter the problems of multiple program influences that can occur before, during, and after the subject program's implementation. This can lead to biased estimates of net savings, either higher or lower than actual.

FIGURE 5.1: Single program evaluation components

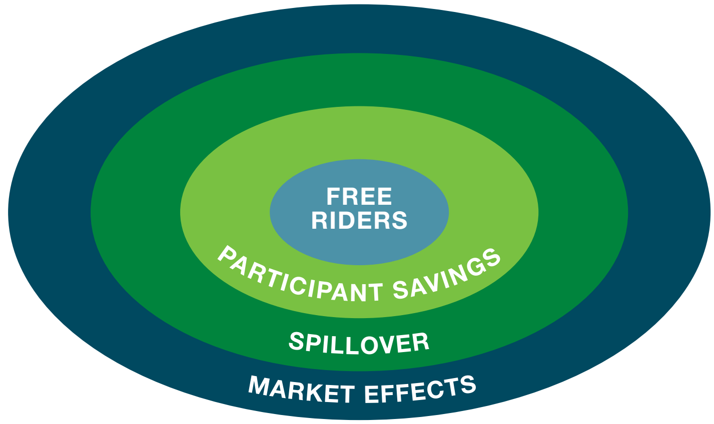
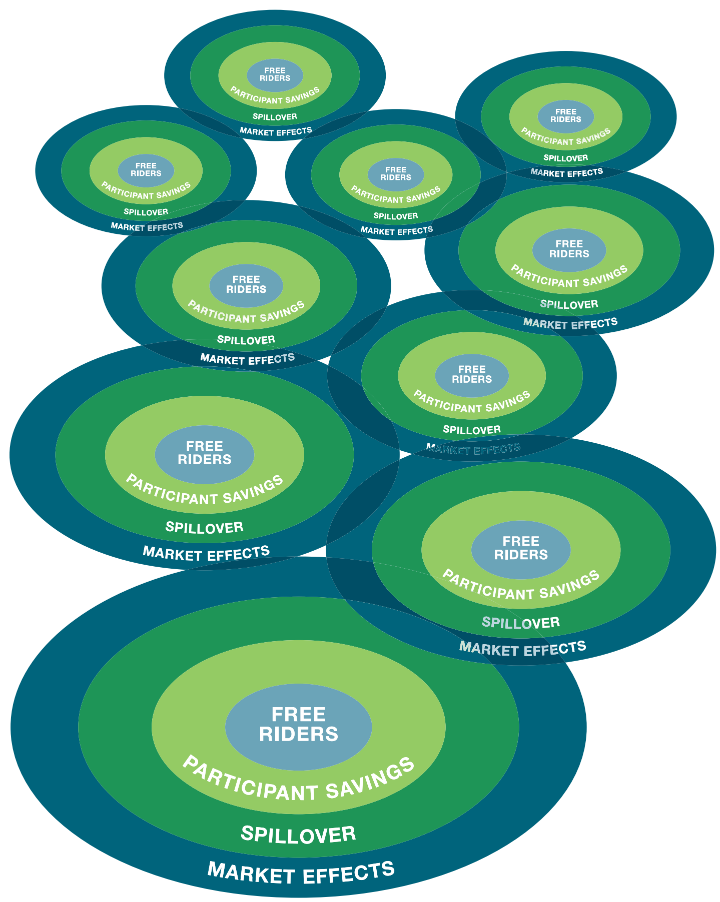


FIGURE 5.2: Multiple program evaluation overlaps



Source: Mahone, D.; Hall, N. (August 2010). "Pursuit of Aggressive Energy Savings Targets—Aligning Program Design and Evaluation." 2010 ACEEE Summer Study Proceedings; August 15-20, 2010, Pacific Grove, California. Washington, D.C.: American Council for an Energy-Efficient Economy. http://neep.org/uploads/EMV%20Forum/Steering%20Committee%20Notes/DMahone_ACEEE_Paper_for_SC_Notes.pdf.

Implications of these complexities are discussed after the following subsections on the uses of net savings and the methods used to determine net savings.

5.1.2 Importance and Uses of Net Savings

Generally speaking, net savings are of most interest for regulated government and utility programs. In these cases, the responsible party (e.g., a city council or utility regulator) wants to know if the use of public or energy consumer-funded programs are actually having an influence, and thus their efficiency investments are "wise." That is, "Are the programs of interest providing incremental benefits, or do the benefits result from some other influences?" For example, in the case of programs funded by utility customers (the most common situation where the issue of free riders comes up), why would one group of consumers subsidize the efficiency actions of other consumers that would have taken the action anyway? Or, as another example, the environmental benefits of energy efficiency programs are usually considered valid only if they are additional to naturally occurring efficiency activities (i.e., based on some version of net savings).

In contrast, there are other situations where gross savings are of the most interest. For example, private sector energy efficiency performance contracts are a case where gross energy savings are the primary concern. Also, from a "bottom line" resource and environmental perspective, it may not be relevant what exactly caused a change in energy consumption, only that it is occurring or will continue to occur. Table 5.1 summarizes different "public" audiences and their current uses of net savings estimates.

From the net savings scoping paper referenced earlier in this section, as well as other sources, there are at least four uses of net savings that do not appear to be controversial, paraphrased below:

- Using net savings, prospectively, for program planning and design (e.g., for setting consumer incentive levels).
- Assessing the degree to which programs cause a reduction in energy use and demand—with net savings as one of numerous measures that should be given serious consideration in the assessment of program success.
- Obtaining insight into how the market is changing and transforming over time by tracking net savings across program years and determining the extent to which free ridership and spillover rates have changed during the period—and potentially using this insight for defining when, and how, to implement a program exit strategy.
- Gaining a better understanding of how the market responds to the program and how to use the information to inform modifications to program design, including how to define eligibility and target marketing.

The following considerations about net savings are important to define for an evaluation effort, but also can be a potential source of controversy due to the different perspectives and objectives of those who will use the evaluation results:

1. Which factors should be included in the definition of net savings (e.g., free ridership, spillover, and/or long-term market effects/market transformation)
2. Whether common methods of determining net savings do or do not properly account for these factors
3. Whether and how net savings can be used in a retrospective manner, in particular for determining attribution and if an administrator has or has not met its savings goals.

These first two issues are discussed after the following subsection on methods used to determine net savings. The last issue above is similar to the issue of how savings estimates are applied, as discussed in Section 8.3.5.

5.2 APPROACHES FOR DETERMINING NET SAVINGS

Before describing the methods used to determine net savings, it is important to understand that beyond just defining what factors are considered when defining net savings, the actual calculation of 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—such as self-motivation, energy prices, and other efficiency programs—that determine participant and non-participant behavior and decisions. With the increasing "push" for energy efficiency by utilities and governments 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 consumer decisions about whether, when, and to what degree to adopt efficiency actions.

As indicated in Chapter 3, net savings evaluation methods can be categorized as follows:

- **Stipulated net-to-gross (NTG) ratios.** These ratios are multiplied by the gross savings to obtain an estimate of net savings and are based on historical studies of similar programs. Sources of stipulations can cover a wide range, from simply using "negotiated guesses", to historical values (perhaps based on prior year NTG studies), to structured expert judgment panels.
- **Self-reporting surveys**. Information is reported by participants and non-participants without independent verification or review. Respondents are simply asked if they would have

TABLE 5.1: Audiences for Net Energy Savings

AUDIENCE	ESTABLISHED OR EMERGING AUDIENCE	USE OF NET SAVINGS ESTIMATES
Energy efficiency program administrators and planners; energy regulators; legislators; advocacy groups	Established (in some areas, legislators and advocacy groups are emerging audiences)	• Assess if program achieved savings goals • Identify strong and weak areas of program design and redesign program accordingly • Apply strong program designs for other products, in other jurisdictions • Adjust payments to/funding of programs based on goal achievement • Determine if the ratepayer/taxpayer funds are being spent cost-effectively and wisely • Define program administrator financial incentives and/or cost recovery levels
Air regulators	Emerging	• Will apply emission factors to energy savings to estimate greenhouse gas and other avoided emission reductions • Assess degree to which efficiency programs have achieved greenhouse gas reduction and other avoided emissions targets

Source: Modified version of information provided in NMR Group, Inc.; Research Into Action, Inc. (November 2010). "Net Savings Scoping Paper." Submitted to Northeast Energy Efficiency Partnerships: Evaluation, Measurement, and Verification Forum. http://neep.org/uploads/EMV%20Forum/EMV%20Products/FINAL%20Net%20Savings%20Scoping%20Paper%2011-13-10.pdf.

undertaken the action promoted by the program on their own without the incentive (free ridership). Then, they are asked whether they had undertaken additional energy efficiency actions (purchased products or made a behavioral change) as a result of their participation in the program (participant spillover). Through non-participant surveys, respondents are asked if they had recently undertaken energy efficiency actions and if those actions were undertaken as a result of the utility program(s) (non-participant spillover).

- **Enhanced self-reporting surveys.** The self-reporting surveys are combined with interviews and documentation review and analysis. The survey instruments themselves include more enhanced batteries of questions (e.g., would you have taken exactly the same action as promoted by the program or would you have undertaken the action at the same time?).
- **Panel of trade allies.** A significant number of trade allies (e.g, contractors, retailers, builders, and installers) are offered monetary compensation for information on their recent (e.g., last 50) projects. Details requested would include manufacturer, efficiency levels, size, price, installation date, installation location (ZIP code), whether the project was eligible for energy program incentives, whether energy incentives were received, whether any other incentives were received (e.g., tax credits or manufacturer rebates), and an assessment of the program's impact on incented and non-incented efficiency actions. Trade allies would include both program participants and non-participants. This approach, while not currently common, can yield reliable information on standard market practices, and—through an ongoing annual update—provide context for tracking ongoing program impacts or market effects.
- **Large-scale consumption data analysis approaches (randomized controlled trial methods and quasi-experimental methods).** Statistical models are used to compare energy and demand patterns of a group of participants and a control group. Where

a control group of non-participants is used, the savings indicated are "net" of free riders and participant spillover.

- **Cross-sectional studies.** These studies consist of comparisons of market share of targeted technologies or behaviors between a baseline area not served by the program and the area served by the program. The main disadvantage of this type of study is difficulty in obtaining quality data. Also, as energy efficiency programs become more prevalent, finding control areas (areas without similar program activities) is becoming exceedingly difficult.
- **Top-down evaluations (or macro-economic models).** Top-down evaluations use state, regional, or national data at the sector level to assess the extent to which markets for energy-efficient products and services have been affected by programs (See Appendix B).

The following are some general notes on these methods:

- The most commonly used methods are the survey-based self-reporting, stipulation, and the large-scale consumption data analysis approaches.
- Net savings values can be the output of (most) large-scale data analysis and top-down evaluations. With the other impact analysis approaches (M&V and deemed savings), the net savings correction is calculated independently.
- All of these methods can address, to one degree of reliability or another, free ridership. However, the ability of the methods to address spillover is limited in terms of how the method is applied and whether non-participants as well as participants are included in the analyses. Market effects can only be analyzed if the studies are conducted over a long period of time, and if such effects are actually "looked for." When selecting a method, it is very important to define what factors are included in the definition of net savings and whether a selected method can actually and reliably address all of the factors.
- With respect to program size and scale, the survey methods, stipulation, and cross-sectional studies can be used with any program, regardless of the number of participants. However, the top-down and large-scale consumption data analysis approaches can only be used with programs that have a large numbers of participants. This is because the models need large amounts of data to provide reliable results.
- In terms of timing of net savings analyses, it is preferable to analyze a long period of time to address spillover effects. Conversely, the free ridership reporting accuracy is probably highest when the inquiry is made as close as possible to the actual energy efficiency action.
- Each of these methods comes with its own data collection and measurement challenges. Particularly when dealing with subjective indicators, such as why someone chose to implement an energy efficiency measure, care should be taken in not only obtaining the needed data but also in reporting it with appropriate indications of the quality and reliability of results obtained from the data.

In terms of accuracy requirements, the 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. Another uncertainty challenge in surveying is the subjective nature of assigning NTG ratios to each participant—their free ridership and participant spillover "score." A participant is a "total" free rider if he or she would have absolutely installed the exact same project at the exact same time, at the same price, even if the program did not exist—and they know that. Assigning NTG ratios to individual participants is more complicated, however, in cases where the participant had multiple reasons for making the project decision, might have installed a different project, or would have installed it in two years if not for the program—or all of the above. Table 5.2 shows an approach that one evaluator used to define full, partial, and deferred free riders.

The following subsections discuss the more common NTG determination methods: surveys and stipulation of the NTG ratios.

5.2.1 Self-Reporting Surveys

Surveys can be complex to design and administer. Respondents' perception and understanding of the questions is absolutely critical to the success of the inquiries. Surveying approaches have become somewhat standard practice with the guidelines developed for Massachusetts and the self-report guidelines developed by the California Public Utilities Commission's Energy Division.[31]

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, regarding potential bias and 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.

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

TABLE 5.2: Example of Assigning Free Ridership Percentages

FEE-RIDERSHIP SCORE	ALREADY ORDERED OR INSTALLED	WOULD HAVE INSTALLED WITHOUT PROGRAM	SAME EFFICIENCY	WOULD HAVE INSTALLED ALL 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

Source: Courtesy of The Cadmus Group, Inc. It is also safe to assign a score of 100% free ridership to those that had already installed the measure prior to receipt of rebate.

in the survey; for example, people who have 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 a NTG analysis that is not very well planned, funded, and/or executed.

Another form of survey bias is *response bias*: the tendency of respondents to gauge their responses to conform to socially acceptable values. This issue is well recognized in social sciences and is discussed in a vast body of academic and professional literature. Another aspect of response bias is construct validity, which raises questions about what the survey results actually measure. The problem stems from the fact that while survey respondents, by virtue of their participation in the program, are predisposed to efficiency, it is not clear to what extent their responses are conditioned by the effects of the program itself.

Generally, the best means for implementing self-reporting surveys have involved asking a series of questions, with each question allowing a scale of responses. 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 confidence and precision in the estimate. Based on the responses to the various questions, each response is assigned a free-ridership score. 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.

5.2.2 Enhanced Self-Reporting Surveys

To improve the quality of NTG ratios 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 include the following:

- **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 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. A common barrier is financial (project costs), so an analysis is done 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.[32] The other type of analysis is to review 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.
- **Market data collection.** Through the review of other information resources prepared for similar programs, the survey data can be adjusted. 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 written or Web-based participant and non-participant self-surveys to triangulate on an estimate of the free ridership and spillover.

5.2.3 Stipulated Net-to-Gross Ratio

This approach, although not a calculation approach, is often used. NTG ratios are stipulated in some jurisdictions when the net savings value is not considered critical, or if the expense of conducting NTG analyses and/or the uncertainty of the potential results are considered significant barriers. In such a situation, a regulatory body sets the value. It is the simplest approach, but one with a high potential for inaccuracy relative to other approaches that involve some level of data collection and analyses.

Sources of stipulated NTG ratios include evaluations of similar programs, hopefully applied to similar populations with a similar level of efficiency adoption and during a time period similar to that of the program being reviewed. Other sources use historical or other information from a wide range of sources to develop a "weight of evidence" conclusion regarding the program's influence. For example, in a three-year portfolio cycle, a stipulated NTG ratio may be used for the second two years based on a NTG ratio determined with other approaches in the first year of the portfolio. One common approach for developing a stipulated value is to use a panel of experts that have relevant technology, infrastructure systems, and market experience.[33] These experts are asked to estimate a baseline market share for a particular energy efficiency measure or behavior and, in some cases, forecast market share with and without the program in place.

5.2.4 Selecting a Net Savings Evaluation Approach and Timing

As mentioned in Chapter 4, selection of an evaluation approach is tied to the objectives of the program being evaluated (e.g., to help understand/improve program design or to adjust savings estimates), 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 one cited for these studies—is the cost of the net savings analysis. The lowest-budget approach is to use stipulated NTG ratios, followed by self-reporting surveys and enhanced surveys, and then various cross-cutting and modeling approaches; although, if the data are available, the top-down evaluation approach can be quite inexpensive. One option for keeping costs down while using the more sophisticated approaches is to conduct an NTG ratio analysis every few years and stipulate NTG ratios for the intervening years as long as the market influences and participants' behavior are relatively consistent.

5.3 ISSUES AND NUANCES OF NET SAVINGS

As noted above, within the energy efficiency industry, it is agreed that net savings determinations are well used for assessing certain programmatic features such as focusing program designs to maximize their efficacy. The net savings controversy is over (1) what factors should be included in the definition of net savings; (2) whether common methods of determining net savings do or do not properly account for free ridership, participant and non-participant spillover, and long-term market effects (i.e., market transformation); and (3) whether administrator goal achievement should be based on net savings or gross savings.

As summarized in the EM&V Forum "Net Savings Scoping Paper" referenced in Table 5.1, the issues can be summarized in one word—attribution:

> Attribution assessment has always involved isolating the effects of the program from other influences. Increasingly, however, when the energy efficiency community mentions the "challenge of attribution" or "sorting out attribution," it refers to the fact that reductions in end users' energy consumption can be affected not only by myriad efficiency programs offered by a broad range of sponsors, but also by economic ups and downs, changes in energy prices, concerns about climate change, and ongoing advances in technology, among other influences. This situation has significantly exacerbated the difficulty of establishing causation, and, therefore, of estimating net savings, and it is likely that this situation will persist. Because of the increased difficulty of establishing causation, some commentators in the energy efficiency community believe that the net savings estimates developed recently are less accurate than those developed in the past when there were fewer programs and messages promoting efficiency and "being green."

The following are some brief discussions covering some specific net savings issues that are subsets of this attribution question. These are presented to not necessarily provide complete answers and recommendations, as the solutions tend to be jurisdiction-specific, but to point out these issues and suggest that these issues be considered as part of the evaluation planning process (see Chapter 8).

5.3.1 Free Ridership

Free-ridership issues are by no means peculiar to energy efficiency; they arise in many policy areas, whenever economic agents are paid an incentive to do what they might have done anyway. However, few issues bring about more discussion in the energy efficiency industry than free ridership. Even the use of the term itself is controversial, as the way in which it is used in the efficiency industry diverges from its classic economic definition.

The basic concept speaks to the prudent use of energy efficiency dollars: they should be spent to encourage customers to take energy efficiency actions that they would not otherwise take on their own. If program dollars are spent on people who would have taken the actions anyway, without program support, then those people are free riders, and those dollars were perhaps misspent. Evaluators are tasked with studying how much of a program's resources were spent on free riders, and what the program savings were, net of free riders. The consequences of free-ridership measurements vary. In some cases, the information is used to refine program plans to better target customers and to assess progress toward market transformation. However, in certain regulatory environments, when free-ridership levels are deemed excessive, program administrators are penalized, claimed savings are discredited, and programs are cancelled.[34]

Beyond the application of free-ridership results, as noted in the section above on spillover and market effects, the actual determination of free ridership can be difficult if the large-scale consumption data analysis approaches of randomized controlled trials and certain quasi-experimental methods are not used (although these methods do not separate out participant spillover and free ridership from one another and do not include non-participant spillover and long-term market effects). And, as noted in the section on methods, the most common method for determination of free ridership is participant surveys—the "self-reports." Self-reports can have problems of response bias beyond the issue of whether people can state their reasons for undertaking efficiency activities (or, for that matter, any behavior that humans undertake).

In addition, in areas with long histories of efficiency programs and activities and many programs operating at the same time, it may not be possible to parse out who is a free rider and who was influenced by the program. In effect, it may be that, in the case of transformed markets or markets being transformed, what is being measured in free-ridership surveys is in fact spillover from other programs.[35]

5.3.2 Spillover and Market Effects

For many energy efficiency programs, the ultimate goal is market transformation, where "interventions" are no longer required for the specific measure or practice because market share is such that adoption is outstripping standard practice. Therefore, it can be a primary evaluation objective for metrics associated with spillover and market effects to be assessed. Unfortunately, many net savings determinations only consider free-ridership levels and/or assess performance for the first year of a program's implementation—too short to assess market effects or long-term spillover benefits. In addition, when assessing free ridership, a survey-based method may be asking the questions of the wrong people. Those identified as free riders might actually be exactly the type of participants that policymakers would want for a market transformation program, those who will take the action and continue to do so once the intervention is over, when the market is transformed.[36] It may be that free ridership is only low in programs (and efficiency actions) that can never result in transformed markets. This may be creating a counter-incentive for administrators to implement programs that cannot result in market transformation—by not counting the savings from their free riders and/or spillover—and giving high marks for programs with low free ridership (i.e., those programs where only an intervention will get people to implement the efficiency activity).

If overly conservative free-ridership measurements are taken, and if free ridership is used to penalize programs, then some program efforts may be killed prematurely before market transformation or ambitious levels of savings are achieved. This also relates to the importance of estimating program spillover and overall market effects. As an efficiency measure is moving up the market transformation "hump," spillover should increase, as there are more satisfied consumers implementing more of the efficiency actions, even without the interventions. Indeed, without the snowball effect of spillover, programs have a large burden to push the measure up and over the hump entirely on their own.[37]

Most current approaches to defining net savings do not address the benefits that programs have on establishing infrastructures that allow future efficiency actions (equivalent to a bus passenger being able to be a free rider only because prior programs resulted in the road and the buses being built and the drivers trained and hired), because program impacts are generally measured over 1–3 years, rather than 5+ years. Net-to-gross ratio adjustments, including those that account for spillover, typically presume that the measure would have been equally available at the same price had no prior programs existed.

The following is an approach to this dilemma, paraphrased from a paper on this subject:[38]

> An approach for policymakers would be to set market targets (e.g., a percentage of market share) and perform market studies that track the progress toward increased market-share. There is no known "tipping point" percent at which any specific technology will be likely to flip to the point of majority adoption (market transformation). But, it is possible to make estimates of the rate of adoption to determine whether the rate of adoption is occurring in a manner that justifies public support, and at what level. This recommendation leaves open the possibility that incentives may be even more important for later adopters than for early ones. This approach allows program design decisions about how to adjust the program over time to be supported by real-time data on market progress and a clear sense of the desired direction for the market through realistic goal setting and adjusting.

With the above points in mind, another conclusion can be that program administrators should budget for tracking of market data, such as sales volumes for specific energy-efficient equipment, price information, and market saturation indicators. This data collection fits under the category of market effects studies. These market effects studies are briefly discussed in Appendix B.

5.3.3 Baselines and Net Savings

Related to determining free-ridership levels and spillover is the point that free-ridership savings essentially raise a question about what the correct baseline is—what would have occurred in the absence of the program or project. In short, would someone have done this anyway? Some baselines, such as a codes and standards baseline or existing conditions baseline, may require a free-rider analysis after the program (project) is implemented to get participant-by-participant indications of free ridership. However, if market-based "common practice" was used to define the baseline for calculating energy savings, the resulting estimates could require no further adjustments for free riders, as the estimates include consideration of what typically would have been done in the absence of the efficiency action. See chapter 7 for more information on baseline selection.

This "common practice" approach to baselines can be used to include consideration of free riders. What "common practice" baselines do not account for is spillover or long-term market effects (i.e., delayed participation/spillover) created by prior programs. However, neither does the use of other definitions of baselines. Therefore, if common practice baselines (i.e., net savings) are to include spillover and other market effects, additional adjustments are required.

5.4 SUMMARY

The above discussion on the factors associated with net savings determination is intended to show the complexities of net savings determination. It is not intended to indicate that net savings should not be determined or that net savings metrics do not have value. As noted previously, there are many valuable uses of net savings determinations.

What is suggested is that the limitations of net savings determination be acknowledged in the policy setting and evaluation planning processes, and that, when net savings are determined, there be clear indications of what factors are and are not included in the determination, and over what time frame. Building on the policy and program implementation strategy recommendation mentioned above, with respect to market effects and spillover, is this overall recommendation:[39]

> Regulators could establish a series of hurdles, or tests, that a program has to pass to avoid high free ridership. The exact nature of the tests would vary depending on the program, but the amount of the incentive relative to the cost of the measure is a good general gauge. When very low incentives appear to attract a large number of participants, or net benefits to

participants are very high, chances are the majority of participants will be free riders. Programs administrators must avoid offering incentives for projects with very short paybacks to participants who most likely would—or should—undertake the project on the project's own financial merits.

[With respect to evaluation activities] program administrators would have to monitor energy-efficient product markets closely to see if a transformation has occurred and exit the market when it has. Expected savings and costs of conservation measures should be revised periodically based on actual saturation of energy-efficient products through well-designed and detailed market effects studies. Baselines can also be adapted over time, based on market baseline studies.

Chapter 5: Notes

[28] NMR Group, Inc.; Research Into Action, Inc. (November 2010). "Net Savings Scoping Paper." Submitted to Northeast Energy Efficiency Partnerships: Evaluation, Measurement, and Verification Forum. http://neep.org/uploads/EMV%20Forum/EMV%20Products/FINAL%20Net%20Savings%20Scoping%20Paper%2011-13-10.pdf.

[29] Kushler, M.; Nowak, S.; Witte, P. (February 2012). *A National Survey of State Policies and Practices for the Evaluation of Ratepayer-Funded Energy Efficiency Programs*. American Council for an Energy-Efficient Economy (ACEEE). Report Number U122. www.aceee.org/research-report/u122.

[30] Mahone, D.; Hall, N. (August 2010). "Pursuit of Aggressive Energy Savings Targets—Aligning Program Design and Evaluation." *2010 ACEEE Summer Study Proceedings*; August 15–20, 2010, Pacific Grove, California. Washington, D.C.: American Council for an Energy-Efficient Economy. http://neep.org/uploads/EMV%20Forum/Steering%20Committee%20Notes/DMahone_ACEEE_Paper_for_SC_Notes.pdf.

[31] Rathbun, P.; Sabo, C.; Zent, B. (2003). *Standardized Methods for Free-Ridership and Spillover Evaluation—Task 5 Final Report* (Revised). Prepared for National Grid, NSTAR Electric, Northeast Utilities, Unitil, and Cape Light Compact. www.cee1.org/eval/db_pdf/297.pdf.

California Public Utilities Commission (CPUC). (2007). *Guidelines for Estimating Net-To-Gross Ratios Using the Self-Report Approaches*. CPUC Energy Division Master Evaluation Contractor Team. ftp://ftp.cpuc.ca.gov/puc/energy/electric/energy+efficiency/ee+workshops/selfreportguidelinesdetailed_v20.pdf.

[32] Simple payback is a common metric but it is complicated to apply. For example, what one entity might consider a short payback another might see as a long payback. Also, because the payback is usually calculated based on energy cost savings, it ignores other benefits to participants, such as increase in value of property, more reliability, comfort, and others, and thus can overestimate the payback period from the participant's perspective.

[33] This is called a Delphi process, which is an interactive forecasting method using a panel of experts.

[34] Mahone, D. (August 2011). "Free-Ridership as a Way to Kill Programs—How Evaluation Policies Can Frustrate Efficiency Goals." Boston: International Energy Program Evaluation Conference. www.iepec.org/2011PapersTOC/papers/054.pdf.

[35] Haeri, H.; Khawaja, M.S. (March 2012). "The Trouble With Freeriders." *Public Utilities Fortnightly.* www.fortnightly.com/fortnightly/2012/03/trouble-freeriders.

[36] Haeri, H.; Khawaja, M.S. (March 2012). "The Trouble With Freeriders." *Public Utilities Fortnightly.* www.fortnightly.com/fortnightly/2012/03/trouble-freeriders.

[37] Mahone, D. (August 2011). "Free-Ridership as a Way to Kill Programs—How Evaluation Policies Can Frustrate Efficiency Goals." Boston: International Energy Program Evaluation Conference. www.iepec.org/2011PapersTOC/papers/054.pdf.

[38] Peters, J.; McRae, M. (August. 2008). "Free-Ridership Measurement Is Out of Sync With Program Logic...or, We've Got the Structure Built, but What's Its Foundation?" *2008 ACEEE Summer Study Proceedings*. August 17-22, 2008, Pacific Grove, California. Washington, DC: American Council for an Energy-Efficient Economy (ACEEE). www.aceee.org/files/proceedings/2008/data/papers/5_491.pdf.

[39] Haeri, H.; Khawaja, M.S. (March 2012). "The Trouble With Freeriders." *Public Utilities Fortnightly.* www.fortnightly.com/fortnightly/2012/03/trouble-freeriders.

Chapter 6

Calculating Avoided Air Emissions

State and federal policymakers and utility regulators are broadening the scope of efficiency programs, and thus their evaluation, by focusing on objectives beyond energy savings. Examples of these broader objectives include reducing the need for on-peak generation (demand response), promoting economic development through job creation, reducing greenhouse gas emissions, and achieving a wide range of air quality and health benefits.[40] Because avoided air emissions in particular are seen as a significant benefit of energy efficiency actions throughout the United States, this chapter has been included to provide guidance for those interested in documenting these benefits.

This chapter first features a summary of the topic of avoided emissions and end-use energy efficiency, which is followed by a discussion of special issues associated with this topic. The remainder of the chapter focuses on ways in which emission factors can be calculated, with one section covering calculation of emission factors associated with avoided onsite fuel use and the following section covering avoided emissions calculations for grid-connected electricity approaches. The second-to-last section provides brief summary comments on selecting a calculation approach, and the last section provides further references on this topic. Also included in this chapter is a discussion of issues related to avoided emissions calculations, including additionality, boundary area definitions, and the design of cap-and-trade programs.

6.1 ENERGY EFFICIENCY AND AVOIDED EMISSIONS SUMMARY

Energy efficiency can reduce air emissions associated with the production of electricity and thermal energy from fossil fuels. Air emissions that can be reduced by efficiency include the six commonly found air pollutants (also known as the U.S. Environmental Protection Agency [EPA] “criteria pollutants”): particle pollution (often referred to as particulate matter or PM), ground-level ozone, carbon monoxide, sulfur dioxide, nitrogen oxides, and lead. All of these pollutants, as well as others such as mercury, have geographic-specific impacts; therefore, where the emission reduction occurs is critical to determining the benefits of efficiency-induced avoided emissions.

The other major potential avoided emissions are greenhouse gases (GHGs)—primarily carbon dioxide (CO_2)—from fossil fuel combustion. Energy efficiency is very important for reducing GHGs because there are few options or “controls” for reducing CO_2 emissions from combustion once the CO_2 is formed, and because, unlike the pollutants mentioned above, the impact of GHG reductions is not location dependent. Therefore, 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 60% of all human-induced (anthropogenic) GHG emissions come from energy-related activities.[41]

Historically, emission reductions from efficiency activities were usually only described subjectively in program evaluations as a non-quantified (non-monetized) benefit. This is changing with increasing interest in quantifying these benefits for at least two purposes:

- Determining the cost-effectiveness of efficiency programs (to possibly justify more investment in efficiency) by monetizing the environmental benefits of efficiency
- Supporting state claim of emissions benefits in state air pollution plans (e.g., State Implementation Plans—SIPS—for criteria pollutants) or GHG reduction requirements (e.g., California’s Assembly Bill 32: Global Warming Solutions Act[42]).

EFFICIENCY AS A COST-EFFECTIVE EMISSIONS REDUCTION STRATEGY

Energy efficiency policies and programs offer the potential to achieve emissions reductions at a cost that can be lower than traditional control measures. The EPA is gaining experience with these potentially cost-effective strategies in rulemakings affecting the utility and other sectors. For example, a recent EPA modeling scenario for EPA’s Mercury and Air Toxics Standard rule predicts that moderate levels of energy demand reduction—equivalent to the continuation of current policies—could lower total compliance costs, reduce ratepayer bills over the long term, and, in some cases, delay or avoid the need for equipment upgrades or new construction of generating facilities and emissions controls. This energy demand reduction is also likely to reduce emissions of air pollutants on high electricity demand days when air quality can be especially harmful.

Source: U.S. Environmental Protection Agency. (2012). The Roadmap for Incorporating Energy Efficiency/Renewable Energy Policies and Programs into State and Tribal Implementation Plans. www.epa.gov/airquality/eere.

The federal government (including EPA and DOE) and many states have recognized the value of incorporating end-use energy efficiency into air regulatory programs. Several programs—including EPA's Acid Rain Program, EPA's NO_X Budget Trading Program, and the Regional Greenhouse Gas Initiative (an effort of nine states from the Northeast and Mid-Atlantic regions)—have provided mechanisms for encouraging energy efficiency.[43] The EPA has also provided guidance on incorporating energy efficiency into SIPs[44] and has approved the inclusion of energy efficiency measures in individual SIPs.[45]

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 (i.e., 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.** This approach involves 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 per year). There are several sources of emission factors, as well as approaches for calculating the factors. The emission factors approach can be used with any project type or energy resource, such as boiler retrofits that save fossil fuels or lighting projects that save electricity. Below is the basic equation for this approach: avoided emissions = (net energy savings) x (emission factor)
2. **Scenario analysis approach.** This approach involves calculating a base case of source (e.g., electricity generating units connected to a grid) emissions without the efficiency programs and comparing that with the emissions of those sources operating with the reduced energy consumption associated with the efficiency programs. This is done with a range of approaches, from using historical generation and load data or capacity factor data with emission rates in spreadsheet calculations to the use of sophisticated computer simulation approaches using "dispatch models." Scenario analysis is typically only used with large-scale, electricity-saving programs. Below is the basic equation for this approach:

   avoided emissions = (base case emissions) – (reporting period emissions)

More information about each of these approaches and their applications is provided later in this chapter.

6.2 SPECIAL CONSIDERATIONS FOR CALCULATING AVOIDED EMISSIONS FROM EFFICIENCY PROGRAMS

This section describes some critical considerations for efficiency program evaluators to consider when determining avoided emissions. One important consideration for both the avoided emissions calculation approaches listed above 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. The following are three potential causes of the difference:

- Different definitions of additionality
- Different definitions of boundary areas
- Characteristics of emissions control mechanisms/regulations that may be in place.

The first two items are discussed in Sections 6.2.1 and 6.2.2, respectively. The "cap-and-trade" emissions control mechanism and its attributes with respect to energy efficiency are discussed in Section 6.2.3. Although it is not the only option to achieve widespread emission reductions, it is addressed here because of its unique characteristics and its current application in the United States for controlling both criteria pollutants and GHG emissions.

Following these sections is a brief overview, in Section 6.2.4, of the possible objectives associated with calculating avoided emissions and how they can affect decisions about which calculation approaches should be used and which specific issues should be addressed.

6.2.1 Additionality

Additionality is the term used in the emissions 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. Note that additionality is typically defined as project-based, versus program- or portfolio-based, and is directly related to the efficiency evaluation issue of defining proper baseline conditions and in some cases to participant free ridership, as described in Chapter 5. As with baselines, additionality 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 formally claiming avoided emissions benefits 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.

One area of common consideration for additionality is whether an efficiency program is mandated by existing legislation (e.g., an energy efficiency resources standard), regulation (e.g., a building code), or policy (e.g., a government agency's or private company's policy to reduce energy use by a certain percentage through efficiency actions). That such legislation, regulation, or policy was influenced by anticipation of an entity being required to meet a future greenhouse gas reduction target or simply to improve air quality would argue for such efforts as being additional. However, that such efforts will occur irrespective of additional support from an emissions program would argue against such an additionality finding. This conflict is illustrative of the policy decisions to be made in deciding additionality.[46]

One approach for addressing the additionality conflict is to review what was assumed in terms of baseline energy use and efficiency activity when a greenhouse gas or air emissions target is set. Levels of efficiency activity included in an air emissions baseline might very well not be considered additional once an air emissions target is set based on such a baseline; that is, the level of emissions already assumes that level of efficiency activity. However, if the efficiency activity is beyond what is included in the baseline, then the program may be considered additional.

6.2.2 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.[47] 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 lower energy use at the efficiency project sites. This is particularly important for calculating avoided emissions associated with electricity efficiency projects, when the affected facilities are connected to the grid. Because power plants are interconnected through the electric grid system, to fully understand electricity-related emissions requires an understanding of the operation of power plants throughout the system, which usually involves regional considerations. See Section 6.4.

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 (e.g., high-efficiency, fossil fuel 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., of an electric generating unit, or EGU). 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:

- **Primary effect:** 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.
- **Secondary effect:** an unintended change in emissions caused by a program. Secondary effects are sometimes called "leakage." Leakage[48] and interactive effects[49] 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 and upstream and downstream effects:
  - **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 potential 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. A scenario for this would be if production at the factory that underwent the efficiency actions is shifted to a factory outside the boundary area or to one in the boundary with less-efficient energy use.

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, it is possible that secondary effects could undermine the primary effect; therefore, the emissions assessment boundary should be defined and leakage possibilities considered, even if only documented that there are no significant, identifiable, secondary effects. One way to document such secondary effects would be to conduct a generic study of the relative upstream (and other) emissions associated with energy-efficient products versus conventional products and apply the results as a "factor" to the calculation of avoided emissions (if such a factor were reliable and relevant).

In summary, when evaluating the avoided reductions associated with efficiency programs, it is important to properly define the assessment boundary and to account for all primary effects (the intended savings), as well as all direct emissions (at the project site) and indirect emissions (at other sites). In addition, ideally, secondary effects should be at least qualitatively investigated.

6.2.3 Special Issues for Capped Pollutants under Cap-and-Trade Programs

There are several regulatory 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 facilities. Allowances are created to represent the emission of each unit (e.g., 1 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 emissions cap or limit on an individual facility, and each facility'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 facility or sell allowances if it has a surplus.

There are several examples of U.S. cap-and-trade programs:

- The Title IV acid rain sulfur dioxide (SO_2) trading program sets a cap on annual SO_2 emissions for U.S. power plants.[50]
- The Clean Air Interstate Rule (CAIR) includes a cap-and-trade mechanism for power plants in the eastern United States to achieve reductions of emissions that cross state lines and contribute to ground-level ozone and fine particle pollution in other states.[51]
- CO_2 emissions from power plants are capped in the nine states of the Northeastern Regional Greenhouse Gas Initiative[52]; in 2011, California enacted a multi-sector cap-and-trade program to limit GHG emissions.[53]

The level of the cap is an important aspect of a cap-and-trade program. In general, emissions may not exceed the cap, and they are also unlikely to be below the cap during any substantial period of time. 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 emit.[54]

The fact that capped emissions tend to remain at the cap level is very relevant to the effect of energy efficiency, which reduces the output of electricity generators or perhaps very large industrial boilers that could be subject to a cap. When emissions are not capped, energy efficiency reduces emissions. As noted, this is not typically true for emissions from sources subject to caps (e.g., large boilers, power plants). This is because reductions in the output of electricity generators do not alter the overall cap on emissions from electricity generators, and any reductions in emissions (and demand for allowances) at a particular generator as a result of energy efficiency make extra allowances available for other entities to use. This means that freed-up 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.

The goal of the cap-and-trade program is typically not to go below the cap but to achieve the cap at the lowest possible cost to society, so energy efficiency contributes to the primary goal of the cap-and-trade program by helping to achieve the emissions target while minimizing compliance costs. In addition, efficiency programs may reduce emissions from non-capped emission sources and non-capped pollutants and directly claim avoided emissions if properly calculated.

There are, however, mechanisms by which efficiency programs under a cap-and-trade system can claim avoided emissions for capped pollutants. The primary mechanism is that allowances are retired, or removed from the market. In states that have established an energy efficiency set-aside program (a pool of allowances from within the cap that is set aside and made available for energy efficiency project and program implementers) as part of their cap-and-trade program design, efficiency program implementers may acquire allowances to retire by submitting a claim for allowances to the set-aside administrator. Alternatively, efficiency programs may acquire allowances to retire by purchasing them from other market participants.

For example, some states have created special set-aside allocations of allowances in their NO_X trading programs for energy efficiency projects.[55] 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. Under CAIR, states have an opportunity to develop energy efficiency (or renewable energy) set-aside programs.

Lastly, it should be noted that while efficiency generally does not reduce aggregate emissions for capped pollutants, efficiency may affect the temporal or geographic distribution of emissions in ways that bring environmental benefits, as well as possibly reduce the on-site emissions footprint of the entity undertaking the efficiency actions. For example, efficiency programs that target cooling loads may reduce electric generating unit NO_x emissions on hot summer days with high ozone concentrations in ways that reduce pollution levels, even though aggregate NO_x emissions for the ozone season would be unchanged. For more information on the air-pollution impacts of energy efficiency under capped systems, visit www.epa.gov/airquality/eere.html.

6.2.4 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 documenting legitimate avoided 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, boundary area definitions, and other issues specific to the program.

However, the following are some rule-of-thumb recommendations, organized by objective, for calculating the avoided emissions:

- **Calculating avoided emissions primarily for informational purposes.** When the primary goal of an efficiency program is to save energy or demand, the avoided emissions are often reported only subjectively or with minimal analysis, to indicate a co-benefit of the program. 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. Thus, one of the simpler approaches described below is probably appropriate. It is typical that (1) additionality is simply assumed, (2) emissions boundary area issues are not necessarily rigorously addressed, and (3) 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.
- **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 or to comply with a regulatory mandate. In these situations, documentation should be provided (either on a project-by-project basis or, preferably, on a program or policy level)[56] that the energy savings (probably net energy savings) and avoided emissions are additional or surplus as defined by the air regulator. A boundary definition is also desirable to document that there is no anticipated emissions "leakage." That boundary definition properly defines the locations of the avoided emission sources—from the efficiency project sites to within a multi-state electric grid. In the case of regulatory mandated air emissions control programs, the methods for calculating avoided emissions may be defined. In terms of the actual emissions-reduction calculations, one of the more complex approaches described below will most likely be appropriate.

6.3 DIRECT ONSITE AVOIDED EMISSIONS CALCULATIONS

Direct onsite avoided emissions can result when efficiency programs save self-generated electricity that would have been produced at a project site or, more typically, when efficiency reduces the need for onsite 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 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.

  A standard resource is the EPA's AP-42, *Compilation of Air Pollutant Emission Factors*.[57] It is the primary compilation of the EPA's emission factor information and contains emission factors and process information for more than 200 air pollution source categories. A standard resource for GHG emissions is the 2006 *IPCC Guidelines for National Greenhouse Gas Inventories*.[58]

- **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 costeffective as a part of a certification process.

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 emissions.

The following is an example of an avoided CO_2 emissions calculation for a project that reduces natural gas consumption from a large industrial boiler.

- First, avoided natural gas use is calculated, in units of therms per year (1 therm = 100,000 Btu):
  - Displaced steam use due to efficiency project = 10,000 million Btu (MMBtu)/year
  - Steam boiler higher heating value (HHV) efficiency = 80%
  - Displaced natural gas use = 10,000 MMBtu/year ÷ 0.80 = 12,500 MMbtu/year = **125,000 therms/year**.

- Next, an emission factor is calculated:
  - The average carbon coefficient of natural gas is 14.47 kilograms (kg) of carbon per MMBtu. The fraction oxidized to CO_2 is 100% (from the Intergovernmental Panel on Climate Change [IPCC]).
  - CO_2 emissions per therm are determined by multiplying the following: natural gas heat content; the carbon (C) coefficient; the fraction oxidized; and the ratio of the molecular weight ratio of carbon dioxide to carbon (44/12). When using this equivalency, keep in mind that it represents the CO_2 equivalency for natural gas *burned as a fuel*, not natural gas *released to the atmosphere*. Direct methane emissions released to the atmosphere (without burning) are about 21 times more powerful than CO_2 in terms of their warming effect on the atmosphere.
  - MMBtu/1 therm x 14.47 kg C/MMBtu x 44 grams (g) CO_2/12 g C x 1 metric ton/1,000 kg = **0.005 metric tons CO_2-equivalent/therm.**

- Finally, annual avoided emissions is calculated:
  - 125,000 therms/year x 0.005 metric tons CO_2/therm = 625 metric tons CO_2 equivalent/year.

This example is provided, in part, to demonstrate that even something as simple as an emission factor calculation can be complex and involve more than just reading an emission factor off a chart. It is important to understand the assumptions and applicability of the factor, such as in the above case for calculating CO_2 emissions (equivalent) from a carbon emissions value. Therefore, it is always suggested that people familiar with the use of such emission factors be consulted when doing this type of analysis.

6.4 EMISSION FACTORS FOR GRID-CONNECTED ELECTRIC GENERATING UNITS

As with the direct onsite emissions cases, emission reductions from reduced electricity consumption occur because less fuel is combusted. However, calculating avoided electrical grid emission reductions is more complex because the fuel combustion in question would most likely have occurred at many different existing or future electric generating units (EGUs) connected to the grid.[59] Thus, emissions from displaced electricity usually depend on the dynamic interaction of the electrical grid, emission characteristics of grid-connected power plants, electrical loads, market factors, fuel and electricity economics, and a variety of regional and environmental regulatory factors—all of which can change over time.

When using electricity savings values for calculating avoided emissions, a critical consideration is to convert the electricity savings at the site when the efficiency measures are implemented to the electricity savings at the EGU. This means taking into account transmission and distribution (T&D) losses between the end use (e.g., home, office, factory) and the generator. The difference between site savings and savings at the generator (sometimes called source or busbar savings) varies with a wide range of factors, with typical values of 5%–20%.

This section of Chapter 6 first provides some basic information about the electric grid and the relationship of efficiency-displaced existing (and perhaps future) electricity generation and avoided emissions. This is followed by descriptions of several approaches, from simple to complex, for calculating avoided emissions.

6.4.1 The Electricity Generation Mix

The electric grid is composed of a T&D system, often covering multiple states, 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 east coast electric utility.

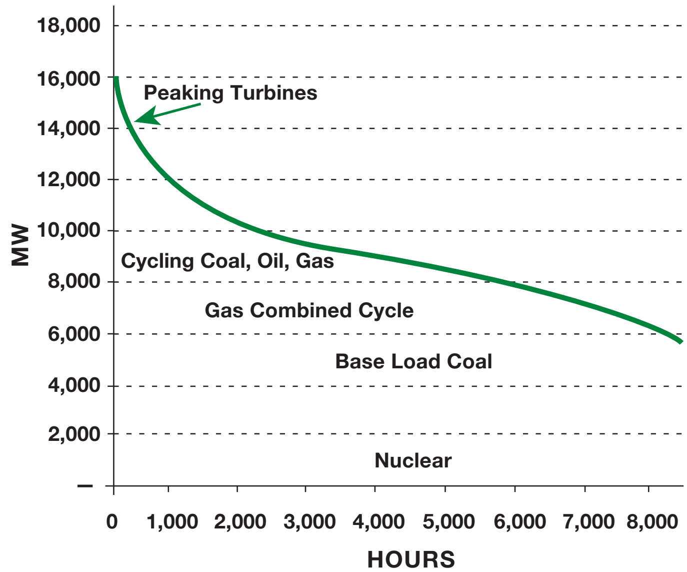
The figure shows that the highest hourly electric demand was 16,000 MW and the lowest was 5,500 MW. It also shows that the peaking turbines and reciprocating engines operated for only about 200 hours per year (in this case, during very hot hours of the summer), while the baseload coal and nuclear plants operated throughout the year. The total area under the curve is the generation needed to meet load plus line losses (in this case, about 80 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 along with operation and maintenance directly related to production and perhaps regulation-based loading orders that prioritize certain resource types, such as renewables. Baseload units are operated as much as possible, unless there is an environmental regulation limitation, because they are the least expensive. On the other hand, peaking and intermediate (cycling) units are used only when needed because of their higher costs. The type of units—such as baseload or peaking—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.

FIGURE 6.1: Example electricity load duration curve



The load duration curve in Figure 6.1 depicts an existing generation mix. However, efficiency could also prevent the need for future power plant construction. 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 one 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 (1) determining appropriate emission factors for the BM and OM using the emission factor approach, or (2) accounting for new and existing generating units when using the scenario approach. This is discussed further in the following subsections.

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

ER = (w) x (BM) + (1 – w) x (OM)

Where:

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

- *w* is the ratio (between 0 and 1) assigned to the build margin
- *BM* is the build margin emission factor (e.g., tons of CO_2-equivalent/MWh)
- *OM* is the operating margin emission factor (e.g., tons of CO_2-equivalent/MWh).

Time is explicit in the above equation. That is, the emission 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 equation, w indicates where the generation produced (or reduced) by the project activity would have come from in the baseline scenario. A ratio (w) 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 ratio 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 ratio 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 World Resources Institute/World Business Council for Sustainable Development (WRI/WBCSD) *Protocol Guidelines for Quantifying GHG Reductions from Grid-Connected Electricity Projects*.[60] In this WRI/WBCSD approach, there are three options for selecting the BM emission factor:[61]

- **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 (1) the baseline candidate (i.e., baseline power plant) with the lowest barriers or greatest net benefits, or (2) 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 and justification for its use as the most likely scenario. 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 thermal 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.

6.4.2 Using Avoided Emission Factors and Scenario Analyses for Electric Grid 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, available resources, data quality requirements, and evaluation framework requirements. Designers of programs or regulations that use these estimates may also wish to specify one or more methods at the outset, and a process for periodic review of those methods.

The emission rates of the electric 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. 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, as well as evaluation framework requirements.

The following are descriptions of four emissions quantification approaches. These four approaches are listed in order of increasing complexity of assumptions and sophistication of analysis:

1. Regional non-baseload emission rates (using EPA's eGRID database)
2. Regional marginal baseload emission rates (using capacity factors or equivalent)
3. Regional historical hourly emission rates
4. Energy scenario modeling.

Table 6.1 provides some guidance on selection of an approach for a given set of objectives and metrics.

TABLE 6.1: Choosing an Avoided Grid Electricity Emissions Quantification Approach

ANALYTICAL QUESTIONS	EMISSIONS QUANTIFICATION APPROACH	ENERGY DATA NEEDS	EMISSIONS OUTPUTS
What is the relative magnitude of emission reductions of the energy efficiency portfolio?	Load emission rates (using EPA's eGRID database)	Annual or seasonal energy impacts (MWh)	Regional non-baseload avoided emissions
Which EGUs in my region are on the margin, and how much seasonal or annual emissions will be avoided?	Regional marginal baseload emission rates (using capacity factors or equivalent)	Annual or seasonal energy impacts (MWh)	Regional marginal unit avoided emissions
How can I quantify hourly emission reductions? How much are emissions reduced during peak electricity demand?	Regional historical hourly emission rates	Hourly energy impacts (MW and/or MWh)	Regional hourly avoided emissions
How will EGU emissions change in future years? How can I compare baseline and forecast emissions that can result from efficiency portfolios? How can I estimate avoided emissions in a cap-and-trade program?	Energy scenario modeling	Hourly, seasonal, or annual energy impacts, which depend on the model used (MW and/or MWh)	Regional avoided hourly, seasonal, or annual emissions based on dispatch order

6.4.2.1 Regional Non-Baseload Emission Rates Approach (Using EPA's eGRID Database)

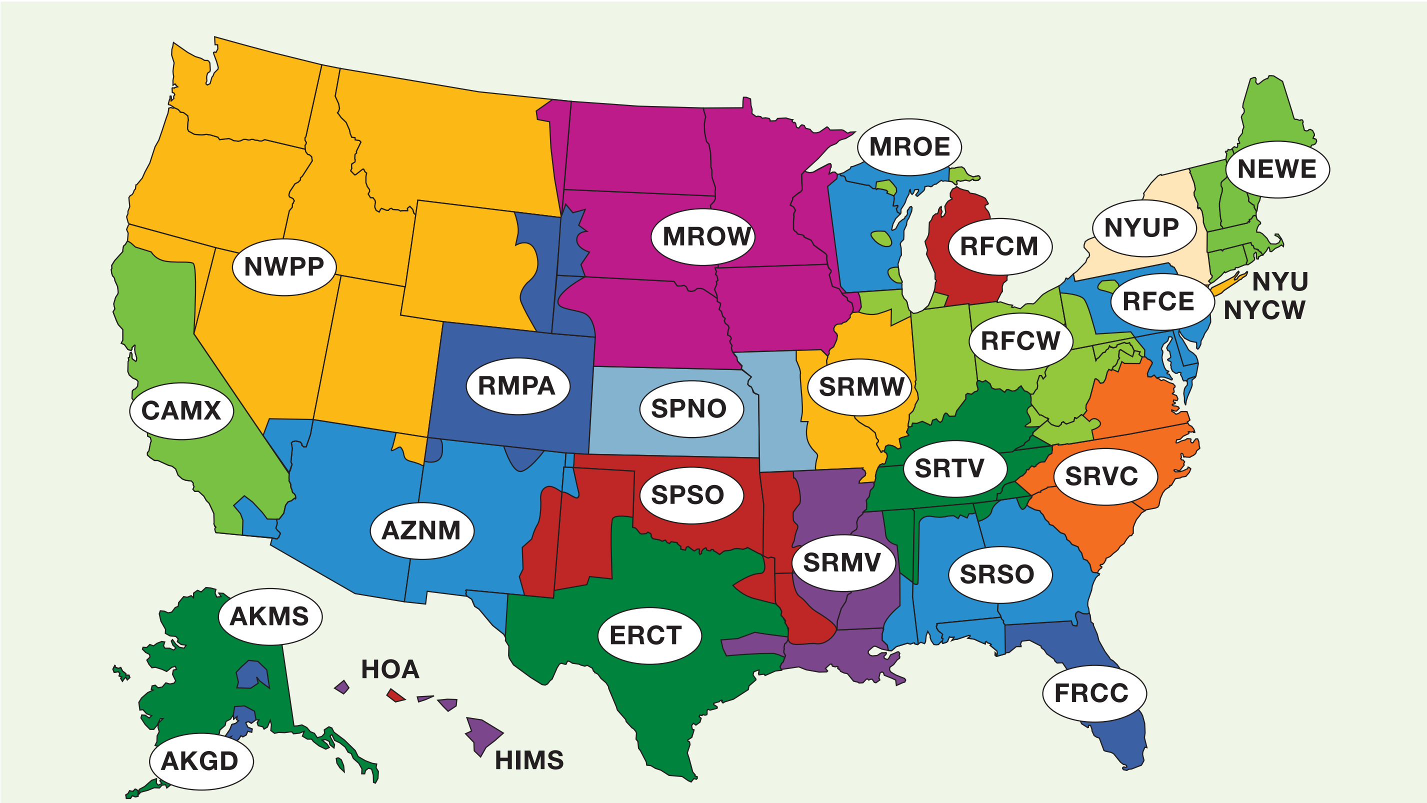
This approach entails a simple calculation—multiply the amount of generation or electricity sales displaced by the efficiency action by the "non-baseload" emission rate indicated for a specific pollutant in a region. The non-baseload emission rate represents an average emission rate for the EGUs that are likely to be displaced by end-use efficiency actions.

A standard source of emission rates for the United States is the EPA's eGRID database. The eGRID database includes operational data such as total annual emissions and emission rates (for GHG, NO_X, SO_2, and mercury), generation, resource mix, capacity factors, and heat input. The eGRID emissions are associated with the generation of electricity, and thus the values do not account for T&D losses, imports and exports among eGRID subregions (or any other geographic area), transmission constraints within any geographic area, or lifecycle emissions at EGUs (e.g., emissions from the extraction, processing, and transportation of fuels).

Figure 6.2 shows the eGRID subregions, which are identified and defined by EPA, using the North American Reliability Corporation (NERC) regions and power control areas as a guide. An eGRID subregion is often, but not always, equivalent to an Integrated Planning Model (IPM)[63] subregion. The 26 eGRID subregions in eGRID2010 are subsets of the NERC regions, as shown in Figure 6.2.

The eGRID subregion non-baseload output emission rates are used to estimate emission reductions of end-use efficiency programs, portfolios, or policies that reduce consumption of grid-supplied electricity. Non-baseload output emission rates are associated with emissions from plants that combust fuel and are the EGUs most likely to back down when energy efficiency and renewable energy, policies, and programs are implemented. These emissions data are derived from plant-level data and are aggregated up to the eGRID subregion level.

FIGURE 6.2: U.S. EPA eGrid subregions



Source: U.S. Environmental Protection Agency. The Emissions & Generation Resource Integrated Database (eGRID). www.epa.gov/egrid/

The main advantages of the eGRID subregion non-baseload emission rates approach are that it is a straightforward, simple calculation and that it can be used for communicating to the public that emission reductions can result from the implementation of efficiency programs. A possible 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 or an exterior lighting retrofit with savings only at night; thus, using an annual average emission factor that lumps daytime, nighttime, weekday, and weekend values together may skew the actual emissions benefits calculation.

A system-average emission rate may also 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 emission rate could be estimated for a hypothetical future system, based on assumptions about emissions from new plants and future regulatory effects on existing plants.

The bottom line is that this is an easy approach to apply, but the tradeoff can be relatively high uncertainty in the estimate of the emissions-related impacts. In summary, the advantages of this approach are as follows:

- It provides an easy "back of the envelope" calculation
- Non-baseload output emission rates provide a basic understanding of how much EGU emissions could likely be avoided or displaced.

And, in summary, this approach also has some limitations:

- Future-looking EGU representation is missing (i.e., there is no "build margin" analysis)
- Some EGUs in the base year may have already shut down or will do so in future years
- The eGRID approach uses averages and does not show where or which EGUs will be displaced
- Information is generally on a 3-year time lag.

- The eGRID approach only accounts for generation within a specific area and does not include information about imports/exports of electricity (except for state-level net imports)
- The approach assumes that efficiency activities will affect all non-baseload plants proportionally to each plant's non-baseload generation.

6.4.2.2 Regional Marginal Baseload Emission Rates Approaches (Using Capacity Factors or Equivalent Approaches)

These approaches have been developed to provide a reasonably accurate estimate of displaced emissions at a lower cost than analyses using the next two approaches, historical hourly data or modeling, but with more accuracy than the simple, non-baseload approach discussed above. Using regional marginal baseload emission rates typically involves using spreadsheets and compiling publicly available data to approximate which marginal generating units will be supplying power at the time that efficiency resources are reducing consumption.

The two major steps in a spreadsheet-based analysis are to (1) determine the relevant set of generating units (i.e., account for the location of the efficiency program's projects as well as transfers between the geographic region of interest and other power areas), and (2) estimate the displaced emissions from those units.

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 can use readily available public data, it is based on a simplified view of the dispatch process that does not account for transmission congestion.

As with the system-average approach, this method does not provide a way to determine how large a geographic region should be considered or how interregional transfer is estimated. However, this approach improves upon the system-average approach with respect to identification of marginal generators. 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 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, having seasonal weekday and weekend and nighttime and daytime values (i.e., six emission factors) to match up the net efficiency savings for the equivalent time period will significantly improve estimates over the other emission factor methods previously described above.

One of these marginal baseline approaches is to use EGU capacity factor data. An EGU's capacity factor is the ratio, in a given period of time, of the actual electricity produced by a generating unit to the electricity that could have been produced at continuous full-power operation. The capacity factor of an EGU can be used as a proxy for how likely the EGU is to be displaced by end-use efficiency actions. The approach helps users understand the relative dispatch order of the EGUs within a state or group of states. After doing this analysis, one has an estimate of which EGUs are on the margin (essentially the load duration curve) and how much emissions could be displaced from each EGU on a seasonal or annual basis.

For example, this approach assumes that EGUs with low capacity factors (e.g., operating at equal to or less than 20% of capacity) are most likely to be displaced by the efficiency portfolio, and EGUs with high capacity factors (e.g., operating at equal to or greater than 80% of maximum capacity) would not be displaced by the efficiency activity. When available, seasonal capacity factors should be used instead of annual capacity factors, which ignore seasonal weather variations. For example, many combustion turbines only operate during summer daytime hours in a typical year. Using an annual capacity factor would incorrectly allocate displaced emissions to these units during seasons when they are not operating.

In summary, the advantages of these approaches are as follows:

- Emissions can be assigned to each EGU
- The calculation is relatively easy if the analysis infrastructure is set up
- It is a simple way to get a relative sense of the marginal unit in the area of analysis.

These approaches also have lmitations:

- Annual capacity factors assume that the EGUs operate the same throughout the year
- Emissions estimates are approximate, based on annual or seasonal capacity factors, and they do not account for maintenance or outages
- Imported and exported power is not considered
- They assume that EGU generation characteristics are the same in the base year and future years

- They assume that all energy savings or generation affect all peaking units first, which is not always true (e.g., street lighting programs).

As an example of this approach, since 1993, ISO New England Inc. (ISO-NE) has annually analyzed the marginal emission rates of the New England electric generation system. This is motivated by the need to determine the emission reductions that demand-side management (DSM) programs have had upon New England's aggregate NO_X, SO_2, and CO_2 generating unit air emissions. The use of these emission rates was subsequently broadened to include the benefits of renewable resource projects in the region. The 2010 report is available at http://iso-ne.com/genrtion_resrcs/reports/emission/final_2010_emissions_report_v2.pdf.

6.4.2.3 Regional Historical Hourly Emission Rates Approach

This approach requires technical manipulation of historical generation, load, and emission rates to determine EGU dispatch order and marginal emissions rates. By applying this approach, one can determine where each EGU fits within the dispatch order for each hour, day, and month of a historical year—and thus determine which EGU's emissions are avoided by less energy demand, due to efficiency, in a given hour. This approach does not account for electricity imports or exports into or out of the grid being studied, nor does it specifically address transmission constraints.

With this approach, one can understand (for every hour or segment of hours of a historical year) which EGUs are baseload (operating all hours of the day), which EGUs are load-following (EGUs that ramp up or down depending upon demand), and which EGUs are peaking units (EGUs that only operate at high demand periods). This approach is most appropriate to answer questions such as: "How much emissions are reduced in blocks of hours, or during periods of peak electricity demand?" or "How much emissions are reduced for demand-response policies?"

The EPA has information that can be used for this approach. It collects generation, emissions, and heat input data in hourly intervals from continuous emissions monitoring systems (CEMS) for all large EGUs subject to EPA trading programs. For example, the EPA implements the emissions cap-and-trade program for the Acid Rain Control Program, the NO_X Budget Trading Program, and the Clean Air Interstate Rule. These programs require an hourly accounting of emissions from each affected unit. *Affected units* are sources participating in CEMS that provide hourly emissions data unless the EGU qualifies to use one of the alternative monitoring methodologies specified in EPA rules.[64]

Other examples of data are available as well:

- The Mid-Atlantic Regional Air Management Association (MARAMA) completed an hourly emissions analysis of the states in the Northeast and Mid-Atlantic region.[65]
- The Metropolitan Washington Council of Governments used a time-matched marginal emissions approach that matches certain efficiency (and renewable) technologies or measures with historical hourly emissions information from the EPA hourly database. This emissions tool can be used for the Virginia/Maryland/Washington, D.C. area.[66]

To use this approach, one needs to identify whether the efficiency affects peak hours and/or baseload energy use. The evaluator should add the programs together in a "bottom up" approach to obtain an aggregate level of energy savings and generation on an hourly basis, and then apply their impacts to the predicted displaced EGUs. The steps associated with this approach are as follows:

1. Collect EGU hourly emissions data from EPA or other resources.
2. Determine EGU dispatch order (and thus hourly emissions) and prepare hourly "bins" of emission rates.
3. Apply hourly end-use energy efficiency savings, with a T&D factor, to the "bins" of emission rates to quantify hourly avoided emissions as a result of the efficiency activities.

The advantages of this approach are as follows:

- Reported data are easy to find on EPA's website on an hourly, daily, and quarterly basis
- Emission rates from any group of hours can be derived from the hourly data.

This approach also has limitations:

- Setting up an hourly emissions database can be resource intensive if the infrastructure is not established
- Representation of future EGU emissions is missing
- Only EGUs subject to EPA's national reporting requirements are represented in its hourly database
- Energy import and export exchanges and transmission constraints are not captured.

6.4.2.4 Energy Scenario Modeling Approach

At the other end of the complexity spectrum from calculating simple average emission factors are the energy scenario modeling approaches that use dynamic simulation models of the grid. These are generically called dispatch models. Varieties or different names for dispatch models include *capacity expansion models, production* *cost models, and system planning models*. Dispatch models forecast which EGUs will operate at any given time based on inputs and assumptions in the model, and their algorithms simulate the complex interactions of the grid with consideration of factors such as transmission constraints, import/export dynamics, fuel prices, air pollution control equipment, and a wide range of energy policies and environmental regulations. Dispatch models specifically replicate least-cost system dispatch, with the lowest-cost resources dispatched first and the highest-cost last. All of these models can capture a high level of detail on the specific EGUs displaced by energy efficiency projects or programs.

The models are used to generate scenarios of the electric grid's operation and emissions. 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 this scenario approach is that a dispatch model is run with and without the efficiency actions, and the resulting difference in emissions is calculated. The models can also be used to provide hourly, monthly, or annual emission factors.

Dispatch modeling can be the most precise means of quantifying avoided emissions (assuming good input assumptions and qualified modelers) 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. As such, it is a preferred approach where feasible.

On the downside, dispatch modeling typically involves the use of proprietary, commercial programs; requires extensive underlying data; and 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 portfolios 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; the California Public Utilities Commission, for example, used dispatch modeling to determine the avoided greenhouse gases from various efficiency portfolios.[67]

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 it can represent an anticipated future system based on detailed assumptions about additions, retirements, and major grid changes (e.g., capacity expansion or system planning models). 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.

Dispatch models using only data for existing units are generally good to use for projecting avoided emissions 1-5 years into the future, especially when the future EGU fleet is not changing substantially. Capacity expansion models forecasting future generation and retirements, as well as the dynamic fluctuation within the electric grid, are generally useful for analysis 5-30 years into the future. This approach is most appropriate to use when a portfolio is expected to be large enough to substantively change electric system operations.

The advantages of this approach are as follows:

- It is the most sophisticated way to capture how the electrical grid will react to implementation of efficiency actions
- Electricity transfers are well represented
- It can provide very detailed estimations about specific plant and plant-type effects
- It can provide highly detailed, geographically specific hourly avoided emissions data at the EGU level.
- Future EGU generation and retirements can be represented
- It can be relatively inexpensive to use for an efficiency portfolio analysis if the models are already developed, and appropriate data populated for other purposes, such as supply-side generation analyses.

This approach also has limitations:

- The models are only as good as the assumptions used
- Hourly emissions data are not always available as the required input
- Energy models are proprietary, require significant resources to run compared to other approaches, and can be data intensive
- Input assumptions can be difficult to discern due to the proprietary nature of the models and the amount and complexity of assumptions and data used
- Expertise in energy modeling is normally recommended.

6.5 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, per common practice, standard emission factors can be used.

For electricity savings programs, system average emission values can be used, but they should be avoided except in the simplest estimates. There are also medium-effort approaches (using capacity factors or historical emissions rates) that can fairly accurately quantify the effects of electricity energy efficiency programs. However, the most sophisticated 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.

6.6 ADDITIONAL RESOURCES ON AVOIDED EMISSIONS 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 www.wri.org/climate.

- *Guidelines for Quantifying GHG Reductions from Grid-Connected Electricity Projects,* www.ghgprotocol.org/files/ghgp/electricity_final.pdf, published in August 2007
- *GHG Protocol Corporate Accounting and Reporting Standard (Corporate Standard),* www.ghgprotocol.org/files/ghgp/public/ghg-protocol-revised.pdf, revised edition, published in March 2004
- *GHG Protocol for Project Accounting (Project Protocol),* www.ghgprotocol.org/files/ghgp/ghg_project_protocol.pdf, published in December 2005.

Examples of energy efficiency projects implemented for their greenhouse gas emissions benefits can be found at the Climate Trust website: www.climatetrust.org.

For criteria pollutants, the following document from the EPA provides guidance on calculating avoided emissions for both energy efficiency and renewables programs: The Roadmap for Incorporating Energy Efficiency/Renewable Energy Policies and Programs into State and Tribal Implementation Plans, available at www.epa.gov/airquality/eere (see Appendix I of the roadmap). Additional EPA references are provided in Appendix C.2.3.

CALCULATION OF AVOIDED EMISSIONS: WISCONSIN'S FOCUS ON ENERGY PROGRAM

Evaluators for Wisconsin's Focus on Energy (Focus) public benefits energy efficiency program estimated emission factors for the plants serving Wisconsin. They used these data to estimate environmental impacts, in the form of displaced power plant emissions, associated with Focus energy savings. 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 the EPA for the power plants supplying Wisconsin (EPA's "Acid Rain Hourly Emissions" data series). The evaluation team aligned its method for estimating emission rates with recommendations of the Greenhouse Gas Protocol initiative (GHG Protocols) developed by the World Resources Institute (WRI) and the World Business Council for Sustainable Development (WBCSD). Emission factors from reduced use of natural gas at the customer sites were also taken from EPA data.

Using the emission rates and evaluation-verified gross electricity savings estimates, the Focus programs together potentially avoided 8.7 million pounds of NO_X; 10.7 million pounds of SO_X; 6.6 billion pounds of CO_2; and more than 41.5 pounds of mercury from inception to December 31, 2010 (see Table 2-23 of the Focus on Energy Evaluation Annual Report [2010], Revised June 17, 2011).

One implication of adherence to the GHG Protocol is that emission factor calculations are based on generation data specific to the geography of efficiency programs. The relevant set of plants from which emissions are displaced are those that serve the electric grid in the areas where the efficiency programs are implemented. A second implication of following the GHG Protocol is that emission factors are estimated only for plants that are operating on the margin, that is, the plants most likely to have remained offline as a result of a reduction in demand/consumption resulting from energy efficiency programs.

To identify marginal plants, the evaluators calculated the average length of time, in hours, that a generating unit remains on once it has been brought online. Peaking units, which are brought on for only a short time, have a short average time on; baseload plants that remain on for hundreds of hours or more have a long average time on. The evaluators divided the population of generating units into five groups: those averaging less than 6 hours on, 6-12 hours on, 12-24 hours on, 24-96 hours on, and more than 96 hours on for each time they are dispatched. Marginal emissions in each hour were defined as those produced by the set of generating units in the group with the shortest average time on. Because the EPA data allow an 8,760-hour accounting of pollutants, insofar as energy savings can be assigned to hours of the day and days of the year, a more accurate emission rate can be estimated by matching the amount of energy savings in a given hour to the emission rate for that hour. Focus evaluators call this approach time of savings (TOS) emission factors.

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 were also subject to a cap-and-trade program for NO_X (the Clean Air Interstate Rule), which has the same effect.

Provided by David Sumi of The Cadmus Group

Chapter 6: Notes

[40] U.S. Environmental Protection Agency. (February 2010). *Assessing the Multiple Benefits of Clean Energy: A Resource for States*. www.epa.gov/statelocalclimate/resources/benefits.html.

[41] Intergovernmental Panel on Climate Change (IPCC). (2007). *Climate Change 2007: Synthesis Report*. www.ipcc.ch/pdf/assessment-report/ar4/syr/ar4_syr.pdf. Contribution of Working Groups I, II, and III to the Fourth Assessment Report of the IPCC. Share of different sectors in total anthropogenic GHG emissions in 2004 in terms of CO_2-eq.

[42] "Global Warming Solutions Act: Assembly Bill 32." (2006). California Air Resources Board (CARB). www.arb.ca.gov/cc/ab32/ab32.htm.

[43] U.S. Environmental Protection Agency. (August 2011). *Federal Implementation Plans: Interstate Transport of Fine Particulate Matter and Ozone and Correction of SIP Approvals.* www.gao.gov/products/C00127. 40 CFR Parts 51, 52, 72, 78, and 97 [EPA–HQ–OAR–2009–0491; FRL–9436–8] RIN 2060–AP50. ACTION: Final rule.

[44] The Clean Air Act requires state and local air pollution control agencies to adopt federally approved control strategies to minimize air pollution. The resulting body of regulations is known as a State Implementation Plan (SIP).

[45] 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*. www.epa.gov/ttn/oarpg/t1/memoranda/ereseerem_gd.pdf. For example, the Metropolitan Washington Council of Governments developed a regional air quality plan for the eight-hour ozone standard for the D.C. Region nonattainment area that included an energy efficiency measure. The plan was adopted by Virginia, Maryland, and the District of Columbia, and the respective ozone SIPs were approved by the U.S. EPA regions in 2007.

[46] While not directly comparable, another term that is used in the context of the Clean Air Act is "surplus." The definition of surplus depends on how the emission reduction will be used, but basically it implies that one cannot double count a program's emission reductions. As a simple example, if a program's impacts are assumed in the baseline then one cannot claim credit for the impacts in a SIP.

[47] World Resources Institute (WRI); World Business Council for Sustainable Development (WBCSD). (December 2005). *GHG Protocol for Project Accounting* (Project Protocol). www.ghgprotocol.org/files/ghgp/ghg_project_protocol.pdf. Source of the "assessment boundary" and "primary/secondary" terminology.

[48] Leakage: In its broadest terms, *leakage* is the concept that an activity or outcome expected to occur and remain within a defined boundary flows outside the boundary, leading to unintended results. In energy efficiency programs, an example of leakage is when a measure is incented by a program (with the associated costs and assumed savings) but is installed outside of the program's jurisdiction. In the context of air regulation, such as a cap-and-trade program, an example of leakage is a shift of electricity generation from sources subject to the cap-and-trade program to higher-emitting sources not subject to the program.

[49] Interactive Effects: The influence of one technology's application on the energy required to operate another application. An example is the reduced heat in a facility as a result of replacing incandescent lights with CFLs, and the resulting need to increase space heating from another source (usually fossil fuel-fired) or to decrease cooling from another source (usually powered with electricity).

[50] "Acid Rain Program." (2012). U.S. Environmental Protection Agency. www.epa.gov/airmarkets/progsregs/arp.

[51] "Clean Air Interstate Rule." (2005). U.S. Environmental Protection Agency. www.epa.gov/cair.

[52] "Regional Greenhouse Gas Initiative." (2012). http://rggi.org/.

[53] "California Greenhouse Gas (GHG) Cap-and-Trade Program." (2012). California Air Resources Board (CARB). www.arb.ca.gov/cc/capandtrade/capandtrade.htm.

[54] There are some cap-and-trade program design features that may lead to exceptions to this general rule, including banking and borrowing of allowances, a safety valve (where additional allowances above the cap level are made available by the government at a known price), and establishing the cap at levels above business-as-usual emissions (so the emission constraint is not binding on emitters). However, as a first approximation, covered facilities will emit approximately at the cap level.

[55] U.S. Environmental Protection Agency. (September 2005). (Draft). *State Set-Aside Programs for Energy Efficiency and Renewable Energy Projects Under the NOx Budget Trading Program: A Review of Programs in Indiana, Maryland, Massachusetts, Missouri, New Jersey, New York, and Ohio*. http://epa.gov/statelocalclimate/documents/pdf/eere_rpt.pdf.

[56] An example of such a policy is the inclusion of an energy efficiency resource standard in a State Implementation Plan (SIP).

[57] "Compilation of Air Pollutant Emission Factors." (2011). U.S. Environmental Protection Agency. EPA AP-42. www.epa.gov/ttn/chief/ap42.

[58] Intergovernmental Panel on Climate Change (IPCC). (2006). *2006 Guidelines for National Greenhouse Gas Inventories*. www.ipcc-nggip.iges.or.jp/public/2006gl/index.html.

[59] The exception would be if all of the electricity serving the site of the energy efficiency activity is generated in a non-grid-connected, stand-alone generation system.

[60] World Resources Institute (WRI); World Business Council for Sustainable Development (WBCSD). (August 2007). *Protocol Guidelines for Quantifying GHG Reductions from Grid-Connected Electricity Projects*. www.wri.org/publication/guidelines-quantifying-ghg-reductions-grid-connected-electricity-projects.

[61] The WRI/WBSCD Protocol calls for "conservative" estimates, which tend to result in underestimating of benefits, versus using a "most-likely" scenario, which can lead to more accurate estimates.

[62] *"eGRID."* (2012). U.S. Environmental Protection Agency. www.epa.gov/cleanenergy/energy-resources/egrid/index.html. eGRID provides a comprehensive source of data on the environmental characteristics of almost all electric power generated in the United States.

[63] The IPM is a multi-regional, dynamic, deterministic linear programming model of the U.S. electric power sector. It provides forecasts of least-cost capacity expansion, electricity dispatch, and emissions control strategies while meeting energy demand and environmental, transmission, dispatch, and reliability constraints.

[64] "Air Markets Program Data." (2012). U.S. Environmental Protection Agency. http://ampd.epa.gov/ampd/. See hourly emissions data from continuous emissions monitoring systems.

[65] Mid-Atlantic Regional Air Management Association (MARAMA). www.marama.org.

[66] "Promoting Air Quality and Climate Protection with Clean Energy and Energy Efficiency in the Metropolitan Washington Area." Metropolitan Washington Council of Governments (MWCOG). www.mwcog.org/environment/air/EERE/default.asp.

[67] California Public Utility Commission (CPUC) selected a team led by Energy and Environmental Economics, Inc. to model the electricity sector's compliance with AB32, California's Global Warming Solutions Act. www.ethree.com/public_projects/cpuc2.php

Chapter 7

Impact Evaluation Considerations

Chapter 7 has two major portions. The first several sections cover issues that often arise in impact evaluations: determining baselines, determining demand savings, calculating persistence of savings, addressing uncertainty of savings estimates, setting evaluation budgets, and establishing evaluation principles and ethics. Because most of this document focuses on using evaluation to determine energy (and demand) savings as well as avoided emissions from conventional efficiency programs, the second portion of this chapter provides brief overviews of the other evaluation objectives: program feedback, resource planning, and calculating non-energy benefits as well as evaluating some relatively unique program types. These other program types focus on residential behaviors, education and training, market transformation, codes and standards, demand response, or greenhouse gas (GHG) mitigation strategies.

7.1 SELECTING BASELINES

A major impact evaluation decision is how to define the baseline. Baselines are the conditions, including energy consumption and demand, which would have occurred without implementation of the subject energy efficiency activity. Baseline conditions are sometimes referred to as *business-as-usual* conditions and are used to calculate project- and program-related savings. Baselines can also include definitions of non-energy metrics that are being evaluated, such as air emissions and jobs.

Baselines represent the counterfactual condition that is estimated to complete an evaluation. With large-scale consumption data analysis approaches, the baseline is defined by the characteristics and energy use of the control group(s) used in the analyses. However, for the noncontrol group-based impact evaluation approaches, such as deemed savings and measurement and verification (M&V), baseline definitions are determined by the type of project being implemented, site-specific issues, and broader, policy-oriented considerations. These considerations usually result in one of three different types of baselines: existing conditions, common practice, or codes and standards. The following is a discussion about selecting appropriate baselines for the deemed savings and M&V impact evaluation approaches. This discussion first defines the three different types of baselines and then illustrates how baselines are typically set for common efficiency project classifications.

7.1.1 Existing Conditions Baselines

As the title implies, *existing conditions baselines* are what is in place (e.g., equipment, controls, procedures) at the project site before the energy efficiency measures are implemented. For example, for a motor replacement program, it is the energy consumption of the motor that was replaced. As another example, for residential behavior-based programs, it is the energy consumption associated with pre-program behaviors. Typically, for deemed savings approaches, the existing conditions are determined generically at the time of program design (and as perhaps defined in a technical reference manual [TRM]).

For M&V approaches, the existing condition is typically determined at the time of measure/project installation. The common means for determining such existing conditions baselines, assuming sufficient information is available, are through (1) an inventory of pre-retrofit equipment site-specific characteristics, including nameplate ratings, and/or (2) an assessment of the existing equipment's or system's energy consumption rates, based on measurements or historical data. Site-specific characteristics include 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 per fixture) of the replaced equipment, and how many hours the lights would have operated.

DEFINING BASELINE DATA COLLECTION REQUIREMENTS

Assessing baseline and baseline adjustment issues in the planning stage is important for determine data collection and budgeting requirements. The goal is to avoid reaching the analysis stage of an evaluation and discovering that critical pieces of baseline information have either not been collected or have been collected with an unreliable level of quality. This situation can be guarded against by providing specific instructions to program administrators prior to program implementation. Planning for data collection is necessary to give administrators notice and justification for collecting data they would not ordinarily collect for managing and tracking program progress.

7.1.2 Codes and Standards Baseline

Energy codes and standards set minimum requirements for energy efficient design; as such, they affect energy use and emissions for the life of a piece of equipment or building. Codes typically refer to energy requirements associated with construction of new buildings and major renovations of existing buildings (e.g., maximum energy use per square feet or minimum insulation requirements). Standards can refer to buildings or efficiency requirements for specific pieces of equipment such air conditioners and motors (e.g., a minimum motor efficiency).

Thus, *codes and standards (C&S) baselines* are the energy consumption associated with buildings or specific pieces of equipment that meet the legal requirements in place, in the location where a project is implemented. For example, for a motor replacement program, the C&S baseline standard might consist of a motor that meets the minimum requirements of the federal Energy Independence and Security Act of 2007 (EISA) for new motors. Three nuances associated with C&S baselines are as follows:

- Not all efficiency actions and not all measures are subject to codes or standards; thus, the application of this baseline definition is limited.
- A variation on a C&S baseline is when a code or standard exists but is not well complied with; in these cases, one of the other baseline definitions might be more applicable (or a combination of definitions).
- Codes and standards tend to change, and thus C&S baselines need to be regularly updated.

7.1.3 Common Practice Baseline

Common practice baselines are estimates of what a typical consumer would have done at the time of the project implementation. Essentially, what is "commonly done" becomes the basis for baseline energy consumption. For example, if the program involves incenting consumers to buy high-efficiency refrigerators that use 20% less energy than the minimum requirements for ENERGY STAR® refrigerators, the common practice baseline would be refrigerators that consumers typically buy. This might be non-ENERGY STAR refrigerators, or ENERGY STAR refrigerators, or, on average, something in between. Common practice is determined by surveys of participants, non-participants, or analysis of market data. The following are three nuances associated with common practice baselines:

- If a common practice baseline is selected, the resulting savings determination probably excludes, on average, savings from free riders.

BASELINES: THE CONSUMER PERSPECTIVE VERSUS THE PUBLIC POLICY PERSPECTIVE

As discussed in this section on baselines, it is typical for replacement of failed equipment and projects for the baseline to be defined as common practice or a codes or standards requirement. Even projects that are early replacement projects may in subsequent years be evaluated as having a common practice or codes and standards baseline. This makes sense from a public policy perspective, to not incent consumers to buy what they would have normally purchased or what they would be required to purchase, to avoid to over-claiming lifetime savings. However, from a consumer perspective, they are looking at savings from a baseline of what they had before the project was implemented, not what "some evaluator says they would have, or should have, done." They in effect want to see the savings as compared to past energy bills, not hypothetical bills. This difference in perspectives is important to acknowledge when different baselines are selected for determining savings from the "policy" perspective versus the "consumer" perspective, as the savings for each perspective may need to be evaluated differently.

- In a situation where an applicable C&S exists, the common practice baseline and the C&S baseline might be the same value—or not, if the C&S are not complied with or are too "easy."
- As with C&S, common practices tend to change over time; thus, these baselines also need to be regularly updated.

7.1.4 Defining Baselines for Specific Program Types

For the purposes of defining baselines, most programs can be categorized as follows:

- Early replacement or retrofit of functional equipment still within its current useful life
- Improvement of existing processes (e.g., reducing the energy consumption per-unit of production in a factory)
- Replacement of functional equipment that is beyond its rated useful life
- Unplanned replacement of failed equipment
- New construction and substantial existing building improvements (tenant improvements)
- Non-equipment based programs (e.g., behavior-based and training programs).

TABLE 7.1: Standard Practices for Selection of Baselines for Common Program Categories

PROGRAM CATEGORY FOR PURPOSES OF BASELINE DETERMINATION	EXISTING CONDITIONS BASELINE	CODES AND STANDARDS BASELINE	COMMON PRACTICE BASELINE
Early replacement or retrofit of functional equipment still within its current useful life Process improvements	X Existing conditions baseline for the remaining life of the replaced equipment or process	X C&S baseline for the time period after the remaining life of the replaced equipment	X Common practice baseline for the time period after the remaining life of the equipment
Replacement of functional equipment beyond its rated useful life		X	X
Unplanned replacement for (of) failed equipment		X	X
New construction and substantial existing building improvements		X	X
Non-equipment based programs (e.g., behavior-based and training programs)			X What people in a control group would be doing in the absence of the program

Table 7.1 summarizes standard industry practice for defining baselines for each of these categories of programs. Note that these are not mandates; each jurisdiction and each program should establish its own baseline scenarios.

As shown in Table 7.1, the early replacement and process improvement programs present their own challenges in determining the appropriate baselines. This is because the issue of likely remaining life of the replaced equipment (or systems or process) becomes critically important. In contrast to the other program categories listed, early replacement/process improvement programs are geared toward replacing existing (lower-efficiency) equipment with energy efficient equipment before the old equipment ceases to function or before it would otherwise be replaced. In these early replacement/process improvement programs, an approach for defining the baseline can actually result in savings being determined based on two different baselines, for two different periods of time:

- The difference between energy use for the old equipment/process replaced (existing conditions baseline) and the new efficiency measure prior to the time the old equipment/process would typically have failed or ceased to be used ("initial savings rate").
- The difference between energy use associated with a C&S or common practice baseline and the new efficiency measure after the time the old equipment or process would have had to have been replaced ("long-term savings rate").

Thus, ideally, early replacement/process improvement programs would be able to take credit for initial savings during the remaining useful lifetime of the replaced equipment/process. This is the approach that evaluators typically refer to as *dynamic baselines* or the "stair step" approach because of the two levels of savings estimated over the two different periods of time. A key nuance of this dynamic baseline is that it assumes data on age of the existing equipment can be gathered and that the time that a process or piece of equipment could be expected to fail and/or simply be ready for replacement can be reliably estimated. Even so, it is a generally recommended approach for early replacement/process improvement programs.

7.2 DETERMINING DEMAND SAVINGS

For efficiency programs, determining energy savings is almost always a goal of impact evaluations. Energy use 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 or megawatt-hours of electricity consumed over a year). Energy savings results may also be reported by time-of-use period, which breaks the year into several periods coinciding with a utility rate schedule. Examples include peak and off-peak periods of the summer and winter seasons.

In addition, a program's electrical demand savings are also often of interest and, for some programs, are a primary goal, such as with demand response (DR) programs. Electrical demand savings are expressed in terms of kilowatts (kW) or megawatts (MW), which indicate rates of consumption—during a specific period of time. Historically, demand savings (specifically, peak demand savings, rather than simple annual average demand savings) have been harder to define and determine than energy savings. This is because determining demand savings requires data collecting and analysis for specific time periods (e.g., data might be required for summer weekdays between noon and 6 p.m.) compared with just needing aggregated monthly energy data. However, with technology advances lowering the cost of meters, the spread of "smart meters," sophisticated wireless sensors, and related software, it is becoming easier to cost-effectively collect the data needed to calculate demand savings.

The first step in determining demand savings is defining the specific metric of interest (i.e., how demand savings are defined). The following are common demand savings definitions that indicate the specific time period of interest for determining demand savings.

- **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. Similarly, average monthly demand savings or average daily demand savings may be determined.
- **Peak demand savings:** several definitions are used; all involve determining the maximum amount of demand reduction during a "peak" period of time, whether that is annual, seasonal, or a specific period such as during summer weekday afternoons or winter peak billing period hours. If peak demand reduction is to be reported as part of an evaluation, the term must be clearly defined with respect to which time period is associated with the reduction.
- **Coincident peak demand savings:** the demand reductions 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 by the efficiency program. Calculating coincident peak demand requires knowing when the utility has its peak (which is not absolutely 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 an energy user, or a group of energy users, to their coincident maximum demand; it is always equal to or greater than 1.0.
- **DR peak demand savings:** the demand reduction associated with a DR program. DR programs reduce a utility customer's electricity demand in response to dispatch instructions or price signals sent to the program participants—a "call" for reductions. Thus, the DR peak demand savings are determined for when there is a "call" for program participants to reduce their energy consumption rate.
- **Forward capacity market demand savings:** (1) the demand reduction proposed ("bid") to an electricity system operator to meet the level of resource commitments the electricity system operator estimates will be needed to meet future peak demand on the system, and (2) the actual demand reduction that occurs once a commitment is made.[68] (See the sidebar on page 70, the Independent System Operator of New England [ISO-NE] program.)

The calculation for demand savings is straightforward—whether for a year, a day, or a specific 15-minute period of time:

Demand Savings = Energy Savings ÷ Time Period of Energy Savings

Each of the impact evaluation approaches described in Chapter 4, to varying degrees of accuracy and with varying degrees of effort, can be used to determine demand savings using the above equation. The "trick," as mentioned above, is to collect the energy savings data for the intervals of interest (i.e., the time period in the above equation). If annual average demand savings is the only metric of interest, 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 (a typical period of demand recording) energy savings data, or estimates, are required for at least those specific time periods of interest.

ISO-NE M&V MANUAL FOR WHOLESALE FORWARD CAPACITY MARKET

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 (FCM). Demand resources eligible to participate in FCM include demand response, emergency generation, distributed generation, load management, and energy efficiency. They 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 the following:

- **M&V methods.** The sponsor must choose from options largely based on the IPMVP options A through D (or equivalent). It should be noted that ISO-NE deviates from IPMVP guidance in particular for Option A, in an attempt to incorporate use of deemed savings values as an acceptable approach for efficiency programs, given that the M&V manual does not explicitly allow for use of deemed savings as a "methodology," as provided in this guidance document). The ISO-NE manual also allows for other M&V techniques to 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 demand resources, 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%, with an 80% confidence level (two-tailed test) 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 (or proxy variables such as voltage, current, temperature, flow rates, and operating hours) must be a true root-mean square (RMS) measurement device with an accuracy of at least ±2 %.
- **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.

The PJM Interconnection subsequently developed and adopted an energy efficiency M&V manual in 2009, building largely on the ISO-NE M&V Manual. These two documents have largely served to inform the development of wholesale M&V standards and retail M&V model business practices for energy efficiency by the North American Energy Standards Board (NAESB).

Provided by Julie E. Michals, Director, Regional EM&V Forum at Northeast Energy Efficiency Partnerships, Inc. For more information, see the ISO-NE and PJM EE M&V manuals, respectively, at www.iso-ne.com/rules_proceds/isone_mnls/index.html and www.pjm.com/committees-and-groups/closed-groups/eetf.aspx.

Ideally, evaluation results would indicate 8,760 hours (or about 35,000 15-minute periods) per year of energy savings data that could be easily translated into demand savings. However, in practice, both primary and secondary methods for determining demand savings are used. Primary methods involve the actual collecting of hourly, 15-minute, or even "continuous" 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 in regression analyses to account for the effects of weather, day type, occupancy, and other pertinent change variables to determine the demand savings caused by the subject program. Of course, hourly demand data can require hourly independent variable data (e.g., weather) for proper regression analysis.

Secondary methods apply load shapes to collected energy consumption data that are only available as averaged or total values on a daily, monthly, or annual basis. Load shapes indicate energy consumption per hour. These load shapes, for whole facilities or by end use (e.g., lighting, cooling, or heating), may be available from studies of related programs in similar markets. One source for the data is the energy savings load shapes, by measure, included in the California Database for Energy Efficiency Resources (DEER).[69]

For example, load shapes might exist for typical weekdays and weekend days in the form of hourly percent of monthly energy use for an office building. In this example, if the lighting system retrofit in a building saved 3,000 kWh in a month, and the load shapes indicate that the lighting system energy use in a building during the hour of interest (say 4 p.m.–5 p.m. on weekdays) is 1% of the monthly total, then it is assumed that 1% of the energy use is saved during that peak hour—30 kWh, or a peak demand reduction of 30 kW. Another example of secondary data is if peak-to-average demand ratios for lighting systems in the building types of interest are available. So, as another example, for this building that saved 3,000 kWh in a month, or an average of about 4.2 kWh per hour, if the lighting peak-to-average demand ratio is 5:1, then the estimated peak demand savings is 21 kW.

These load shape approaches can be relatively effective for measures with constant performance, such as a lighting retrofit that reduces energy consumption as a percentage of baseline energy use (i.e., change-out of 200 W fixtures for 100 W fixtures). However, these approaches do not work as well with projects with variable savings rates, such as those for a lighting controls/daylighting program or a program that involves installation of variable-speed motor drives on building ventilation fans. In the latter case, there might be zero (or even negative) savings when the fan is at full speed (which often occurs during peak period hours).

7.3 PERSISTENCE OF SAVINGS

One important evaluation issue is how long energy savings are expected to last (persist) once an energy efficiency activity has taken place. While energy and demand savings are often reported with respect to just first-year savings, the longer a measure, project, or program provides savings, the more valuable it is. In addition, a project with a 5-year payback will not be of much value if the efficiency measures only last three years. Thus, estimates of total or lifetime savings are important for determining the cost effectiveness of an efficiency action, as well as their value as a long-term resource.

7.3.1 Definitions

A persistence study measures changes in program impacts over time. Related to these studies are measure-retention studies that assess (1) the length of time the measure(s) installed during the program year is maintained in operating condition, and (2) the extent to which there has been a significant reduction in the effectiveness of the measure(s). Two related terms and their definitions are helpful to know:

- **Measure persistence**: the duration of an energy consuming measure, taking into account business turnover, early retirement of installed equipment, technical degradation factors(s), and other reasons measures might be removed or discontinued.
- **Effective useful life (EUL):** an estimate of the duration of savings from a measure. It is estimated through various means, including the median number of years that the efficiency measures installed under a program are still in place and operable. It is also sometimes defined as the date at which 50% of installed units are still in place and operational. EULs are also defined as "measure life," which is essentially the length of time that a measure is expected to be functional.

Paraphrasing from a recent conference paper,[70] more than 100 energy efficiency persistence-type studies have been conducted, examining in-situ median lifetimes for residential and non residential measures. A review of results from measure-based EUL studies around North America showed that measure lifetimes for the type of measures normally offered in prescriptive programs (programs with defined measures and incentives for implementing the measures) and installed in typical end uses are fairly consistent for many measure-based programs in commercial, residential, and industrial sectors.[71] A key factor affecting the quantity of savings being delivered from program-related installations of equipment is whether the measures perform at the new efficiencies consistently over time, or whether their efficiency performance degrades over time. Decays in net technical performance can be an important issue, particularly for measures for which savings accrue over long periods of time.

As discussed throughout this guide, energy savings achieved over time are a "difference" from a baseline performance. Thus, persistence looks at degradation patterns that would be realized in standard efficiency equipment, or typical (non-efficiency) consumer behaviors, and compares them with the degradation patterns of the program's efficient equipment or behaviors. Therefore, savings are the difference over time between the energy use of the efficient equipment or behavior and the standard equipment or behavior it replaced—with consideration of both baseline and project equipment/behavior degradation in performance (which may be the same).

7.3.2 Rebound

Although it is not usually included in persistence determinations, the *rebound effect* (also called *take back, snap back, back fire,* and a few other terms) resurfaces as an issue periodically, as does its potentially negative impact on energy savings proponents. A form of this concept was first put forth by British economist William Jevons in 1865. An interpretation of the "Jevons paradox" is that increases in efficiency of coal processes would cause coal consumption to increase, to a level that would exceed previous consumption levels. Jevons paradox was based on the industrial sector, where coal costs where a high percentage of operation, and he did not consider the counterfactual, or what would happen if the efficiency of processes did not increase. In any case, Jevons, fundamental argument has been repackaged and repurposed, correctly or incorrectly, to cover micro-level and macro-level impacts; it has been extended to questions about energy efficiency (and carbon emissions reduction) efforts in general, and been given the general term "rebound." In this guide, the *rebound effect* is defined as a change in energy-using behavior that yields an increased level of service accompanied by an increase in energy use that occurs as a result of taking an energy efficiency action.

Few, if any, in the efficiency industry argue that the rebound effect never occurs, although few rigorous studies of rebound have been completed.[72] Clearly, in some situations, an energy efficiency action results in an increased level of service and an associated increase in absolute energy use greater than would be experienced if the level of service did not increase. This is most probable in situations where the pre-project level of service is very low, for example, in low-income housing where prior to the efficiency actions, heating and lighting systems were barely functional, but the new systems allowed the occupants to adequately light their homes and heat them to comfortable levels.

The issue for impact evaluation is whether rebound is explicitly or implicitly included in the savings determination. An example of an explicit consideration is the use of a deemed rebound effect factor, a form of a non technical degradation factor. As with all deemed factors, it should be specific to the applications associated with the subject program and based on actual historical data. Another explicit approach would be a long-term study of rebound in the participants and a control group of non-participants. Current standard energy efficiency evaluation practices do not use either of these approaches or any other explicit approaches for assessing rebound.

What is used commonly, though, is an implicit approach, and although it is not often acknowledged as including rebound effects, it does. The implicit approach is associated with a savings determination method that includes savings determination based on both the project period (after the efficiency project is implemented) and pre-project service levels. This is true with the large-scale consumption data analysis approaches discussed in Chapter 4. If the program participants increase their service level and energy use compared with a control group of non-participants, the calculations will capture this in the savings determination.

There is no "correct" answer to whether or not rebound should be evaluated, particularly given that the magnitude of the rebound effect is not known; however, the evaluation planning efforts and the reported results should indicate whether rebound is addressed or not.

7.3.3 Determining Persistence

Persistence studies can be expensive, and are thus not often undertaken for individual programs. Past experience indicates that long periods of time are needed for persistence studies, and there are significant challenges to conducting these studies, such as long lifetimes of measures (making it impractical to wait for measure failures or consistent patterns of degradation), incomplete data sets, high cost of data collection, and of course, the need for trained staff.

The most accurate manner to report savings persistence is probably as an annual percentage (e.g., 100% in year 1, 98% in year 2, 95% in year 3, and so on). However, the most common way that persistence is indicated is via a value for EUL (e.g., 20 years for an energy-efficient refrigerator).

The basic approaches for assessing persistence are as follows:

- Use of historical and documented persistence data, such as manufacturer's studies or studies done by industry organizations such as the American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE)
- Laboratory and field testing of the performance of energy-efficient and baseline equipment
- Field inspections, over multiple years, of efficiency activities that constitute the subject programs
- 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 2006 California Evaluation Protocols contain a complete section on persistence analyses and can be used to learn more about this subject.[73]

There are two specific nuances of persistence analyses:

- Savings can live as long as the behavior that enables the efficiency is continued. This can be well past the life of the first purchase of that type of technology, or the opposite—well less than the technical life of a measure, if it is not being used. Thus, it is advisable to evaluate persistence of measures associated behaviors (e.g., continuing to use a new control system or always buying the efficient product when replacement is required or responding to information programs) rather than just the effective useful life of a technology.
- As discussed in Section 7.1 on baselines, one baseline option is to use dynamic baselines for early replacement programs. The classic example is a motor with five years of remaining life being replaced with a new efficient motor. For the first five years of the new motor's effective useful life, the baseline is the motor that was replaced. However, for the remaining years of the EUL, the baseline would be a common practice or code/standard compliant motor, as of year five—when it is assumed that the original motor would have been replaced. Some may ask whether this correction in year five is a persistence correction or a baseline correction. Whichever way this may be classified, baseline or persistence, it is important to transparently report measure savings over the lifetime of the motor (and program) with an indication that a dynamic baseline was used if such is the case.

7.4 CONTROLLING UNCERTAINTY[74]

As discussed in Chapter 3, a significant challenge in evaluating energy efficiency programs is the impossibility of direct measurement of the primary metric—energy (and demand) savings. Consequently, the difference between (1) actual energy consumption and (2) what energy consumption would have been if the efficiency measures had not been implemented is an estimate of energy (and demand) savings. While 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 understood and described. While additional investment in the estimation process can reduce uncertainty, trade-offs between evaluation costs and reductions in uncertainty are inevitably required; thus, improved accuracy (and associated EM&V costs) should be justified by the value of the improved information.

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 refers to 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 estimated values relative to the true values (which are not known). As defined this way, uncertainty is an overall indicator of how well a calculated or measured value represents a true value.

Uncertainty of savings-level estimates is the result of two types of errors:

- **Systematic errors:** errors that are subject to decisions and procedures developed by the evaluator and are not subject to chance, also called bias.
- **Random errors:** errors occurring by "chance." One important type of random error is the changes in energy use that can be due to unobserved influences (i.e., unobservable independent variables such as different preferences among people within a population). Additionally, whenever a sample is selected to represent the population—whether the sample is of appliances, meters, accounts, individuals, households, premises, or organizations—there will be some amount of random sampling error. Any selected sample 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 a particular sample. This variability due to chance factors (the "luck of the draw") is termed random sampling error. Random error can be reduced by increasing the sample size.

Because uncertainty arises from many different sources, it is usually difficult to identify and quantify the effect of all potential sources. The distinction, described above, between systematic and random sources of error is important because different procedures are required to identify and mitigate each. Evaluation reports often identify only uncertainty arising from random error, because this source of error is usually the easiest to quantify using confidence intervals and statistical significance tests that are available to provide quantitative estimates of uncertainty caused by random error (see sidebar on statistical terms).

On the other hand, uncertainty attributable to systematic errors does not have a single comparable, quantitative measure. Rather, sources of systematic error/bias are specific to individual studies, depending on equipment used, research staff, or research and data collection procedures employed. To assess uncertainty from systematic sources/bias, it is necessary to address the rigor of data collection, measurements, and analysis. Thus, in summary, uncertainty is typically calculated and reported through the objective analysis of random errors and the subjective analysis of systematic errors.

7.4.1 Systematic Errors/Bias

Systematic errors are problematic because they cause a savings estimate to be biased. Bias is simply the extent to which a measurement or analytic method systematically underestimates or overestimates a value. Systematic errors potentially occur from the way data are measured and/or collected, and/or the way in which analysis (and/or modeling) is conducted. Some specific examples of sources of bias are below.

1. **Measurements**. At times, equipment used to measure consumption may not be completely accurate. Human errors (e.g., errors in recording data) may also cause this type of error. Measurement error is reduced by investing in more accurate measurement technology and training and by more accurately recording and checking data. In terms of specific instruments, the potential magnitude of this type of error is 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.
2. **Data collection.** If some parts of a population are not included in the sample, non-coverage errors result. This may cause a problem because the value calculated from the sample will not accurately represent the entire population of interest.
3. **Self selection.** Other sources of bias include using a control group that is not comparable with the treatment group of participants. For example, in programs where participation is voluntary, a potential source of bias is what is referred to as "self-selection." Participants, by simply choosing to be in the program, are different than the rest of the population. As such, it may be very difficult to find an appropriate control group to use; thus, evaluation findings may potentially produce biased results. Other forms of self-selection bias include those who respond to surveys versus those who do not.
4. **Analysis (modeling).** Estimates are often created through statistical models. Some are fairly simple and straightforward (e.g., estimating the mean) while others are fairly complicated (e.g., estimating response to temperature through regression models). Regardless, errors may occur due to using the wrong model, creating the wrong counterfactual, assuming inappropriate functional forms, including irrelevant information, or excluding relevant information. For example, in modeling energy use of air conditioners, the evaluator may only use cooling degree days as the independent variable if home type or square footage is not available. Thus, the statistical model will attribute all the observed differences in energy use to temperature, while clearly a portion of the use is attributable to the home size. This model will introduce systematic error.

For the measurement-related sources of bias, the best solutions are sound engineering and analysis practices such as calibration of meters, use of experienced and trained personnel; use of rigorous data analysis, and development and rigorous application of quality assurance and control procedures. With respect to bias associated with data collection and analysis, implementing the proper random selection approaches is very important for avoiding bias and mitigating systematic errors in a savings estimate.

7.4.2 Random Errors

There are many possible reasons why energy use might change for any program participant's facility in any particular month (e.g., a household may take a vacation, a commercial office may add employees, a factory might have higher or lower production or change product mix, a school might be in a location with an unusually cold month). An evaluation of a program's impact tries to tease out only the changes in energy use that are due to the program, as opposed to the other factors. While some of the other factors may be easily observable (e.g., occupancy or weather), many factors (e.g., vacations) are not.

Therefore, when change in energy use is estimated, there is some chance (hopefully, a very high chance) that the actual program impact is being measured. However, there is also some chance that other, unobserved factors happened to cause the change in energy use. There is therefore some chance that the evaluation erroneously estimates that the program caused the change in energy use when in fact it was caused by other factors (or combinations of the program and other factors). The same is the case when indicators other than energy use are being determined as a precursor to determining energy savings (e.g., changes in operating hours of lighting systems that have had occupancy sensors installed).

One specific source of random errors that is often a concern is that evaluators typically do not have access to an entire population of interest, 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 population sample data, resulting in sampling errors (one of the major sources of random errors). Note that even if the entire population of interest (e.g., every household in a program) is observed, there can also be random error due to unobserved variables.

The following are examples of energy efficiency program impact evaluation samples:

- **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, or other similarities.
- **Commercial building lighting retrofit program.** A sample of the "spaces" (e.g., offices, hallways, common areas) is selected for determining average operating hours of the lights in each space.
- **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).
- **Net savings analysis of participants in an efficiency program.** A sample of participants and a sample of non-participants are selected for interviews.

Random error as a whole can be estimated by using the laws of probability. In other words, the potential magnitude of the random error for any value calculated from a sample can (usually) be estimated. Sample size can be a particularly important aspect of an evaluation design, and decisions about the sample size are one of the key influences on the random error and the overall uncertainty of the evaluation. A larger sample size will increase the statistical significance of the estimate, which will reduce the chance of the random error.

The common factors for reporting random error-associated uncertainty are confidence and precision. Precision provides convenient shorthand for expressing the interval believed to contain the actual value. The confidence level is the probability that the interval actually contains the target quantity. For example, if the savings estimate is 530 kWh, and the relative precision level is 10%, then the confidence interval is 530 ±53 kWh).[75]

Put another way, if a project with a 10% (of baseline energy use) savings estimate with a 20% precision at the 90% confidence level means that the true program energy savings is between 8% and 12% of baseline energy use, with 90% probability. *It is very important to note that this indication of savings certainty is based on analysis of random error and does not address systematic errors.*

One way to interpret confidence intervals and precision is as a measure of risk. For example, with an estimate of 10% savings, there is some risk that the true savings are more or less than 10%; indications of confidence and precision can help quantify that risk. In general, high levels of confidence that the estimated value falls within the predicted interval can be achieved with wider (less precise) intervals, while narrower (more precise) intervals permit less confidence. In other words, when all else is held constant, there is a trade-off between precision and confidence.[76] As a result, any statement of precision without a corresponding confidence level is incomplete.

For example, suppose the average savings among participants is estimated to be 1,000 kWh per year, and the analyst determines this estimate to have 16% relative precision at the 90% confidence level. The same data set and the same formulas may be used to estimate 10% relative precision at the 70% confidence level. Without reporting the confidence level, the second uncertainty expression seems to have less uncertainty when, in reality, they are identical. In reporting estimates from a sample, it is essential to provide both the precision and its corresponding confidence level (typically 80% to 90% for energy efficiency evaluations).

Before concluding this section on controlling uncertainty, two other terms are introduced: *internal validity* and *external validity*. The above discussion is addressing internal validity—that is, an evaluation is internally valid if the observed results are known to have been caused by the program as opposed to other factors that may have influenced the outcome (i.e., the estimate is unbiased and precise). On the other hand, an evaluation is externally valid if the observed outcomes can be generalized and applied to the population from which the sample was selected, as well as new populations, circumstances, and future years. The external validity of an evaluation is an important consideration if the evaluation results are to be applied to other programs (without their own potential studies or evaluations) operating in different years, locations, and/or populations; for example, external validity answers the question of whether the results of a 2010 Oregon CFL evaluation can be applied to a 2012 version of the Oregon CFL program.

In conclusion, 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, and are the evaluation activities themselves cost effective?

7.5 EVALUATION BUDGETS: BALANCING THE VALUE OF EVALUATION RESULTS AND UNCERTAINTY

This section provides input on establishing a budget for an impact evaluation. Establishing a budget (i.e., setting the funding level) for an evaluation requires consideration of all of the aspects of the evaluation process, particularly consideration of 13 of the 14 issues raised in Chapter 8 on planning (the last issue is evaluation scale and budget itself). This section, however, discusses budgeting in the context of managing the risks of savings uncertainty. It also provides some information on the budgets that different states have assigned to evaluation activities, as an indication of the range of typical budgets.

7.5.1 Using Impact Evaluations to Manage Risk

The evaluation process must find the right balance between interest for the most accurate estimates possible with the costs for obtaining such accurate estimates. This leads to a basic impact evaluation question: "How good is good enough?" Asking "How good is good enough?" is a short version of asking "How certain does one have to be of the energy savings estimate that results from evaluation activities, and is that level of certainty properly balanced against the amount of effort (e.g., resources, time, money) it takes to obtain that level of certainty?" The implication is that energy efficiency investments should be cost effective and evaluation investments

STATISTICAL TERMS

Census (total enumeration) consists of collecting data from each and every unit in the population (e.g., metering every replaced light bulb). Sampling only chooses a small part of the units from the population for data collection and analysis.

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% of observations are within one standard deviation of the mean. 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 that 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 affected 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 that the true average is within ±1% of the estimate (precision), but only be 30% confident that is the case. Alternatively, one may be 99% confident that the true average savings are within ±50% 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 may lead to precision levels that are too large (e.g., more than ±100%) 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%, true average savings are somewhere between a negative 500 kWh (which means the measure actually caused consumption to increase) and 1,500 kWh.

To formalize these relationships, evaluators use a test called the t statistic. The t statistic is a measure of how reliable a statistical estimate is. A t test produces a critical value; for example, at a 90% level of confidence, the critical value of t is 1.645. When the parameter estimate, such as the mean kWh savings, is small relative to its associated variability, the t statistic value is low. When using a 90% confidence level, if the t statistic is less than 1.645, the evaluator concludes the estimated value (e.g., mean kWh savings) is not reliable: it could be negative, positive, or zero.

should consider risk management principles, and thus balance the costs and value of information derived from evaluation (i.e., evaluation should also be cost effective). *Impact evaluation is about managing risk.*

Conceptual approaches that draw upon risk management techniques provide a useful structure for addressing evaluation issues identified in this guide. Unfortunately for energy efficiency, risk management is hampered by the large number of difficult-to-quantify aspects of efficiency and evaluation, although the tools for addressing these difficulties are improving. Energy supply-side resources have uncertainty and risks as well (e.g., uncertainties associated with future fuel costs). However, perhaps the single most identifiable risk of efficiency is the inability to directly measure savings, which creates uncertainty.

Tolerance for uncertainty is driven by how much risk is associated with getting the wrong answer. For example, with energy efficiency, the risks include crediting too much or too little savings to the actions that have been taken to comply with an energy efficiency program objective or mandate, such as an energy efficiency resource standard. This can lead to expending too many resources on ineffective actions (or the opposite), or simply not obtaining the desired outcome (i.e., less energy consumption). However, there is another counterbalancing risk if policymakers minimize energy efficiency as an eligible resource because of this measurement risk. This could take the form of spending too much on EM&V beyond the need for lowering uncertainty and, more important, trying to eliminate this measurement risk by excluding viable (or promising but difficult to "prove") efficiency resources that are replaced with supply-side resources that have different, and perhaps greater, risks associated with their performance and/or lifecycle costs.

An important aspect of risk management is always the relative risk (the "As compared to what?" question), which for efficiency includes comparing the measurement risk with the risks associated with supply-side resources, such as future estimates of fuel costs. Thus, another risk to be managed with efficiency evaluation is to avoid requirements that over-specify certainty and/or the use of overly conservative versus most-likely assumptions, which can result in unnecessarily excluding or limiting energy efficiency as an energy resource strategy.

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."

7.5.2 Budget-Setting Considerations

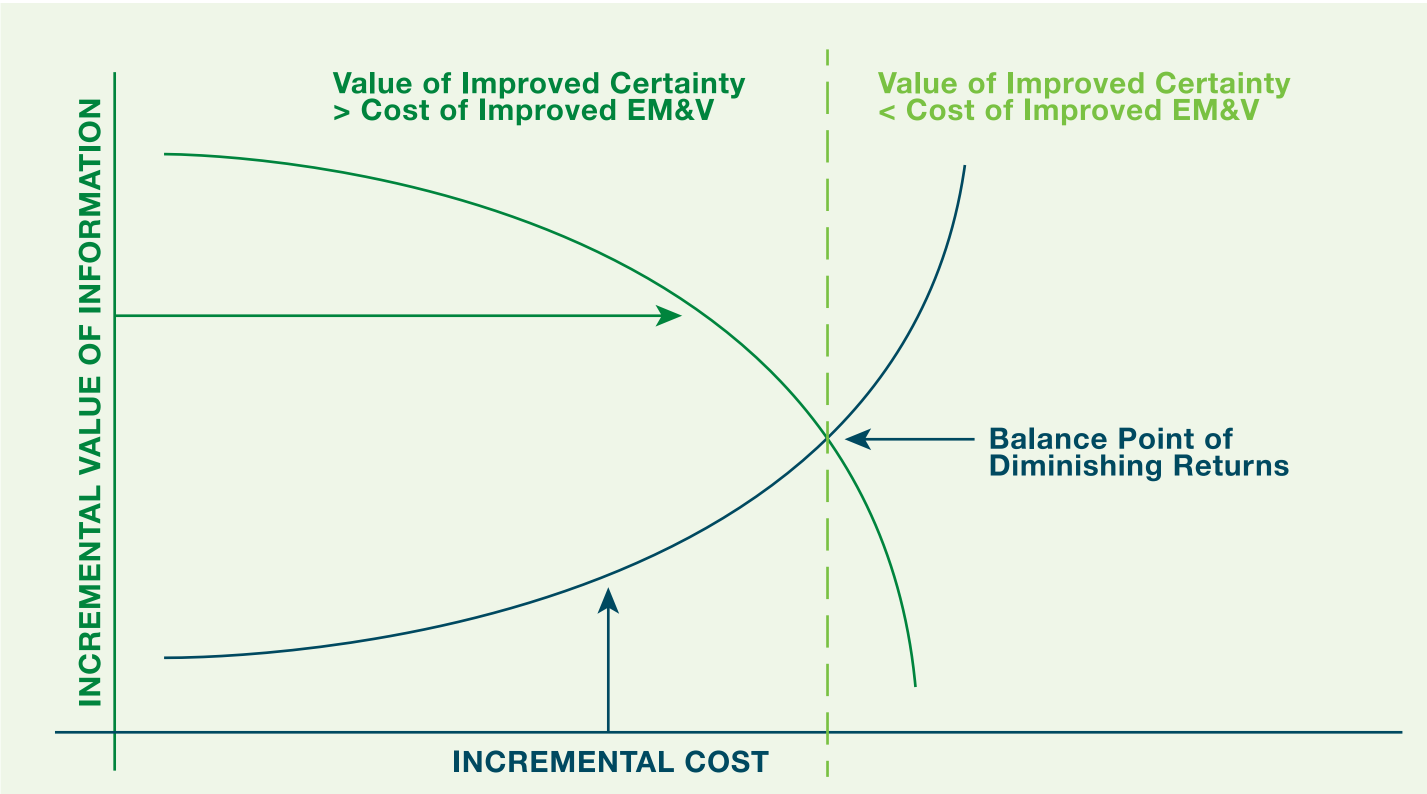
With respect to setting an evaluation budget, the primary challenges are typically balancing (1) the cost, time, and effort to plan and complete the evaluation(s); (2) the uncertainty of various impact evaluation approaches; and (3) the value of the information generated by the efforts. Conceptually, this is shown in Figure 7.1, where the goal is to find the balance point between increasing incremental investments in evaluation (costs) and decreasing incremental value in the evaluation information.

Most of the value of impact evaluation information is tied to the value of energy savings and overall program integrity. In general, 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.

Thus, an important aspect of evaluation planning is deciding what level of risk is acceptable and determining the requirements for accuracy and a corresponding budget. How much risk is acceptable is usually related to many factors, which raise the following questions:

- How large is the program in terms of budget and savings goals? Larger programs tend to have larger evaluations but smaller evaluation costs as a percentage.
- What is the level of uncertainty associated with the expected savings of a program, and what is the risk that the program poses in the context of achieving (or not) portfolio-savings goals (e.g., are the subject program's projected savings 10% or 90% of the portfolio savings goal)?
- Does the savings determination need to indicate how much energy was saved, or just that the savings were above a certain level? The latter is usually easier to determine than the former.
- 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, projected savings estimates. If a fair amount of effort has gone into feasibility studies and perhaps pre-testing (e.g., pilots), then less of an evaluation effort may be required.
- Is it adequate to simply verify that the individual projects in a program were installed (and perhaps operating correctly)? Or, on the other end of the cost spectrum, are rigorous field inspections, data collection, and analyses on all, or a large sample of, projects in a program required?

FIGURE 7.1: Incremental value of information versus incremental cost of evaluation



- Is the program likely to be expanded or contracted? A program that may be expanded (i.e., increased in goals and budget) probably deserves more analyses to confirm whether it should be expanded than one that is likely to receive no additional funding or n be cancelled.
- How long has it been since the last evaluation of the subject program, and has the program changed in the interim?
- Do savings need to be attributed to specific projects within a program (e.g., a custom retrofit program where each project is unique)? If savings values for each project are desired, then a census evaluation is probably required. This is more costly than evaluating a sample of projects.
- How long, in months or years, does the evaluation need to be conducted? Longer evaluation cycles require more funding. On the other hand, evaluations that have started late, and might be in a start-up mode, will require smaller budgets (and probably deliver less detailed results).
- 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 earlier in this chapter, 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 (e.g., a regulatory body) can influence the scale of the effort and budget.
- 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 non-energy benefits to be evaluated and quantified? If this is more than an anecdotal exercise, then additional resources will be required.
- Are the savings reported to be used as “information only” or is there a regulatory or statutory requirement associated with the use of the impact evaluation results? For example:

- Is there a savings goal, perhaps associated with an energy efficiency resource standard, for which the impact evaluation will determine compliance?
- Are there cost-recovery and/or lost revenue recovery financial impacts for utilities to be decided upon by a regulatory body using the impact evaluation results?
- Does the program administrator have a performance-based incentive, the monetary value of which will be determined on the basis of the impact evaluation reported metrics?

Budgets may also be influenced (increased) to accommodate 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. The determination of the appropriate sample size can be a major factor in setting an evaluation budget. To address this, statistical analyses help evaluators determine the sample size needed to ensure that the desired level of precision and confidence for key metrics, or factors that determine key metrics, will be statistically significant. Additional resources (more samples) might also be allocated to ensure that "hard-to-reach" portions of the population are included in the sample (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). DOE's Uniform Methods Project includes information on steps that can be taken to increase the accuracy of evaluation results with respect to sampling.

While it is difficult to generalize, common practice suggests that a reasonable spending range for evaluation (impact, process, and market) is 3% to 6% of a portfolio budget. However, this should be considered rough guidance, because evaluation needs and the relative EM&V roles of program administrators and independent third-party evaluators (and thus how the budget is categorized between program and evaluation expenses) vary significantly between different states and different program administrators. 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 are directly proportional to uncertainty of predicted savings (i.e., projects with greater uncertainty in the predicted savings warrant higher EM&V costs).

7.5.3 State Evaluation Budgets

To provide some guidance on overall budgets for evaluation, the following three tables provide information on the EM&V budgets of various states and program administrators. Table 7.2, from the Consortium for Energy Efficiency's (CEE's) 2011 annual industry report, shows total EM&V spending for program administrators who responded to their survey in the United States and Canada. Looked at in total, for natural gas and electricity programs, the average percentage of program budget spent on EM&V in 2011 was about 3.6%. This is consistent with the range of EM&V budgets for both program administrator efforts and independent third-party evaluator efforts that are cited in Tables 7.3 and 7.4.

Table 7.3, from a 2010 survey report of evaluation practices and issues in the United States, shows total EM&V 2008 and 2009 budgets for a sample of states. Table 7.4, from the same survey report, indicates additional information on EM&V budgets allocated among three main types of EM&V activities (process, impact evaluations, and market research).

Based on information provided by survey respondents, the range of EM&V budgets varies significantly between states from very little to about 6%. The 2008 allocation of EM&V budget among different types of evaluation activities also varied substantially across jurisdictions. Six states allocated 75% or more of their EM&V budget for impact evaluation, while three states and the Northwest Energy Efficiency Alliance (NEEA) allocated between 50% and 60% of their EM&V budget to impact evaluations. It appears that many states allocate between 10% and 20% of their EM&V budget to market research.

Note that these budget figures should be considered as only rough guidance, as they are mostly self-reported, and the definitions that are used for what is and is not included in the EM&V budgets varies significantly between states and program administrators. This is particularly true when looking at where the line is drawn between activities associated with program implementation and those associated with independent evaluation.

7.6 EVALUATION PRINCIPLES

Reliable evaluation of energy efficiency programs requires transparency and independence. This results in high-quality information on which business, regulatory, and policy decisions can be made. Therefore, evaluation processes can be defined by the following principles:

- **Integral to the portfolio cycle.** The evaluation process should be integral to what is typically a cyclic planning-implementation-evaluation process. Therefore, evaluation planning should be part of the program planning process, so that the evaluation effort can not only support program implementation, including the alignment of implementation and evaluation budgets and schedules, but also provide evaluation results in a timely manner to support existing and future programs. See Chapter 8 for more information on the portfolio cycle and the integral role of evaluation.

TABLE 7.2: Electric and Gas EM&V Expenditure and Budget Dollars, United States and Canada (millions USD)*

ELECTRIC			
COUNTRY	2010 EM&V EXPENDITURES	2011 EM&V BUDGETS	TOTAL 2011 ENERGY EFFICIENCY BUDGETS**
United States	58	154	4,239
Canada	11	32	895
Total	69	186	5,134
GAS			
COUNTRY	2010 EM&V EXPENDITURES	2011 EM&V BUDGETS	TOTAL 2011 ENERGY EFFICIENCY BUDGETS**
United States	9	27	782
Canada	1	Less than 1	78
Total	10	27	860

Source: Wallace, P.; Forster, H.J. (2012). State of the Efficiency Program Industry Budgets, Expenditures, and Impacts 2011. Consortium for Energy Efficiency. www.cee1.org/files/2011 CEE Annual Industry Report.pdf.

*Notes: *The above table includes only those programs that provided an EM&V dollar figure in response to the survey. Those that provided an estimated percentage of their EM&V activities from their total energy efficiency funding are not included. **Dollar figures in the Total 2011 Energy Efficiency Budgets column exclude load management because CEE did not ask for EM&V expenditures and budgets in the load management portion of the survey.*

TABLE 7.3: Ratepayer-Funded Energy Efficiency Budgets and EM&V Expenditures for Selected States

STATE/ REGION	TOTAL ENERGY EFFICIENCY BUDGET (MILLION $)	2008 EM&V BUDGET (MILLION $)	2008 EM&V BUDGET PERCENT OF TOTAL ENERGY EFFICIENCY BUDGET	2009 TOTAL ENERGY EFFICIENCY BUDGET (MILLION $)	2009 EM&V BUDGET (MILLION $)	2009 EM&V BUDGET PERCENT OF TOTAL ENERGY EFFICIENCY BUDGET
CA	1014.2	80.2	7.9%	1376.7	79.3	5.8%
CT	113.6	1.7	1.5%	107.4	1.4	1.3%
FL	124.3	?	NA	138.9	0.1	0.1%
IA	58.9	3	5.1%	90.5	3.2	3.5%
ID	19.7	0.9	4.6%	33.1	0.5	1.5%
IL	41.0	1.3	3.2%	67.4	2.2	3.3%
MA	148.9	5.1	3.4%	208.5	7.8	3.7%
ME	16.8	0.2	1.2%	20.8	0.2	1.0%
MN	136.5	1	0.7%	73.7	1.3	1.8%
NEEA	97.5	1	1.0%	105.2	1.8	1.7%
NY	287.9	7.7	2.7%	421.2	7.6	1.8%
OR	76.8	1.6	2.1%	105.4	2.2	2.1%
PA	?	?	NA	8.7	0.1	1.1%
TX	106.4	01	0.9%	101.8	0.2	0.2%
WI	140	2.4	1.7%	162.4	4.8	3.0%

Source: Messenger, M.; Bharvirkar, R.; Golemboski, B.; Goldman, C.A.; Schiller, S.R. (April 2010). Review of Evaluation, Measurement and Verification Approaches Used to Estimate the Load Impacts and Effectiveness of Energy Efficiency Programs. Lawrence Berkeley National Laboratory. Report LBNL-3277E. http://emp.lbl.gov/sites/all/files/lbnl-3277e.pdf. Source of data in table as indicated in the original report: Consortium for Energy Efficiency (2008), Consortium for Energy Efficiency (2009), and U.S. Census Bureau. See source report for notes.

TABLE 7.4: EM&V 2008 Funding Levels and Allocation Among Activities

STATE/REGION	FUNDING ALLOCATION OF EM&V BUDGET (% IMPACT / % PROCESS/ % MARKET RESEARCH)
CA	75 / 15 / 10
CT	NA
FL	NA
IA	50 / 30 / 20
ID	NA
IL	75 / 15 / 10
MA	75 / 15 / 10
ME	30 / 50 /20
MN	NA
NEEA	60 / 30 / 10
NY	80 / 10 / 10
OR	50 / 30 / 20
PA	50 / 30 / 20
TX	75 / 15 / 10
WI***	100 / 0 / 0

Source: Messenger, M.; Bharvirkar, R.; Golemboski, B.; Goldman, C.A.; Schiller, S.R. (April 2010). Review of Evaluation, Measurement and Verification Approaches Used to Estimate the Load Impacts and Effectiveness of Energy Efficiency Programs. Lawrence Berkeley National Laboratory. Report LBNL-3277E. http://emp.lbl.gov/sites/all/files/lbnl-3277e.pdf.

** The range depicts answers provided by different respondents from California; the funding also includes evaluations activities for codes and standards. NA = not available.*

*** Annual funding for the 2009–2013 cycle.*

****Program administrators may conduct some market research during program implementation phase; market research is not carried out as part of EM&V activities.*

- **Useful retrospective analyses.** The evaluation process is designed to support the policy goals of the energy efficiency programs being evaluated. As such, evaluations should develop retrospective estimates of energy savings attributable to a program in a manner that is defensible in proceedings conducted to ensure that energy efficiency funds are properly and effectively spent. Evaluation activities should go beyond documenting savings to actually improving programs and providing a basis for future savings estimates.
- **Adequate resources.** Evaluation budgets and resources should be adequate to support the evaluation goals and the level of quality (certainty) expected in the evaluation results over the entire time frame that program impacts need to be assessed.
- **Completeness and transparency.** Results and calculations should be coherently and completely compiled. Calculations should be well documented and transparent, with reported levels of uncertainty. The scope of the documentation should take into account the relevant independent variables that determine benefits and include a properly defined baseline. In addition, documentation and reporting should include all relevant information in a coherent and factual manner that allows reviewers to judge data quality and results. Among the key qualities of a good, transparent analysis are the following:
  - Project descriptions indicate the approaches and the variables used to determine 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, the uncertainty of their magnitude, and the risk of overestimated or underestimated savings levels. Benefits are calculated at a level of uncertainty such that the savings are neither intentionally overestimated 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 (e.g., utility resource procurement programs or emissions cap-and-trade systems), energy and demand savings

and avoided emissions calculated from one program area 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.

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 what is seen by some as a natural tendency toward savings inflation. There are many real-world incentives for people to over-report savings, 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, many in the evaluation field believe that credibility, not conservativeness, is the desired characteristic, and that underestimates can be just as biased and damaging as overestimates. Thus, the correct guidance is to develop the "most likely" result and not one that is biased to be conservative or aggressive.

Beyond the characteristics of the evaluation itself, evaluations can only be effective if those conducting the evaluations perform their tasks fully and completely, and are free of bias without a stake in the outcome, with respect to the performance of the programs under consideration. 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.

The relationship between the evaluator and the implementers and/or administrators—whose work is being evaluated—needs to be cooperative. This allows for information sharing, access to project sites, and for the results of the evaluator to be considered valid by the implementers and administrators and thus considered as useful input for program improvement. However, there will always be some stress in the relationship as (1) the evaluator cannot allow itself to be unduly influenced by the implementer/administrator, or for that matter, whoever hires the evaluator, including an entity such as a state regulator; and (2) the administrator/implementer will have a sense that their work is being judged by the evaluator, because the evaluator may very well have a significant say in the compensation or penalties applied to the implementers and administrators.

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 website at 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.

7.7 USING IMPACT EVALUATIONS FOR PROGRAM FEEDBACK

Impact evaluation results are used to make informed decisions on program improvements and future program designs and offerings throughout the efficiency portfolio implementation cycle. The implementation cycle is one in which programs are designed, then implemented, and then evaluated. Using the results of the evaluation, programs are reexamined for design changes that may be needed. This cycle provides a continuing process of program improvement, so that the programs match available market opportunities and continually improve their cost-effectiveness over time. The impact evaluation planning process is discussed more in Chapter 8.

Impact evaluations tend to be a retrospective process for determining how a program performed over a specific period of time (e.g., month, season, year); nonetheless, 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 timely and systematic, 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 (e.g., residential, commercial, industrial) basis and feasibility studies tend to be focused on specific customers that may be involved in a particular program.

RESOURCE PLANNING: NORTHWEST POWER AND CONSERVATION COUNCIL

Since 1980, the Northwest Power and Conservation Council has been charged with preparing regional integrated resource plans for the Pacific Northwest states of Idaho, Oregon, and Washington, as well as the western portion of Montana. The council plans consist of forecasts of electric loads and a least-cost/risk-resource portfolio to meet those regional electric demands over a 20-year time frame. All of the council's plans have relied heavily on energy efficiency. Therefore, the council's resource-planning process is critically dependent on data derived from market characterization research and impact evaluation results. Results of these studies support load forecast model calibration and assessments of energy efficiency potential in the following ways:

- The council uses an end-use econometric load forecast model. In simple terms, this model calculates future electricity sales by multiplying the energy use of a given device (e.g., clothes washer, refrigerator) by the expected number of those devices in the region. Market characterization studies, such as residential and commercial appliance and equipment surveys, are used to estimate the saturation of each end-use device included in the forecast model. In addition, these surveys are also used to establish the baseline characteristics of the existing stock, such as the average efficiency of equipment and the lighting power density of residences and commercial buildings. In addition to "savings," impact evaluation results, especially those that provide estimates of both pre- and post-measure adoption energy use, are used to calibrate the forecasting model to actual observed use.
- The council forecast model's estimates of end-use energy consumption and efficiency are used as the "baseline" inputs for its assessment of the remaining potential for energy efficiency in the region. For example, the forecasting model uses the sales-weighted average efficiency and energy use of new clothes washers derived from market assessments and program impact evaluations to compute the regional energy demand created by this appliance. The baseline use and efficiency for clothes washers in the council's energy efficiency potentials assessment is the same as that used in the load forecast. Thus, for resource-planning purposes, the forecast of future demand and the potential for reducing that demand start with the same value. One implication of this approach for impact evaluation is that the use in the forecast and potentials assessment of a "market average efficiency" (even if it is more efficient than local code or federal standards) reflects the efficiency choices that are occurring absent future program effects. As a result, at least in theory, the only difference between gross and net savings is potential spillover impacts.
- The council's assessment of energy efficiency potential and cost (which is also sometimes based on program evaluation data) is used for resource planning. Because the council treats energy efficiency as a resource, it competes directly against generating resources in the agency's resource planning model. That is, rather than reduce the load forecast based on estimates of achievable potential, the amount of energy efficiency to be acquired is based on testing its cost and availability against supply-side resources. Thus, energy savings from impact evaluations are used directly to derive the future need for supply-side resources.

Provided by Tom Eckman of the Northwest Power and Conservation Council

7.8 EVALUATION FOR DEMAND FORECASTING AND ENERGY RESOURCE PLANNING

Efficiency is an important energy resource that is being recognized as such in state and regional energy resource forecasting and planning efforts.[77] This is driven in part by at least 24 states having enacted long-term (three-plus years) specific energy-savings targets, which can take the form of a statewide energy efficiency resource standard (EERS).[78] These EERS targets include goals calling for efficiency to reduce electricity consumption on the order of 1% to 3% per year and/or meet 30% to 100% of increases in demand for electricity. Looking to the future, a 2012 Lawrence Berkeley National Laboratory report found the following:[79]

- Total ratepayer-funded energy efficiency program spending in the United States (80% of which is targeted to electric end uses) is projected to rise from $5.4 billion in 2010 to $7.0 billion in the low scenario by 2025, $10.8 billion in the medium scenario, and $16.8 billion in the high scenario for both electric and gas efficiency programs.
- While the West and Northeast combine for more than 70% of efficiency program spending in 2010, by 2025, it is projected that 50% of the spending will be in the South and Midwest.
- At a national level, the analysis suggests that savings from energy efficiency programs in the medium scenario have the potential to offset up to 80% of the projected growth in U.S. electric load in 2020.

An effect of these actions is that efficiency is, and will increasingly become, a resource that has to be accounted for in state and regional energy-demand forecasts and related capacity, transmission, and distribution resource planning efforts.

Market and impact evaluations play an important role in supporting these resource planning and forecasting activities. Market potential studies can be used to project savings that will be available from future efficiency efforts and impact evaluations, and market effects evaluations can be used to calibrate demand forecast models.

RESOURCE PLANNING: ENERGY TRUST OF OREGON

Energy Trust of Oregon (ETO) impact evaluations, process evaluations, and estimates of market effects are used directly for integration of efficiency into utility resource planning. ETO provides energy efficiency savings under contract, and develops efficiency supply curves and deployment scenarios in Oregon for PacifiCorp, Portland General Electric, Northwest Natural, and Cascade Natural Gas. ETO also provides the same services for Northwest Natural in their Washington territory. This occurs through the following steps:

1. ETO develops supply curves—estimates of the cost and savings available from various conservation measures and sectors. For each integrated resource planning (IRP) cycle, estimates of costs are updated as needed based on ETO program data. The estimates of savings are adjusted based on the most recent available evaluations.
2. ETO develops deployment scenarios that provide a recommended rate of acquisition over 20 years. Impact and process evaluations inform that rate to the extent that they show where and how ETO's programs are reaching customers. These scenarios show "net" savings after market effects, based on trends from recent evaluations.
3. Utilities integrate efficiency into their linear planning models for resource selection (as part of the integrated planning process) in a variety of ways. The common factor is that the measures that are less costly than generation in the long term are selected. This is affected by the savings estimates.
4. ETO funding to pursue future conservation is determined based on goals coming out of this IRP process. Success is then judged by evaluation estimates of net savings.

Provided by Fred Gordon of Energy Trust of Oregon

Impact studies that consider savings attribution can also be used to transparently indicate the sources of changes in demand forecasts (whether from voluntary efficiency programs, codes and standards, economic conditions, changes in social norms, or other factors). Cost-effectiveness analyses can be used in integrated resource planning efforts for comparing efficiency with other energy resources such as conventional and renewable power plants, as well as location-specific transmission and distribution requirements.

For evaluations to support demand forecast and resource planning efforts, it is important for the evaluation planning efforts to consider the metrics and data requirements, as well as time frames of the people conducting the resource forecasting and planning efforts. These people include forecasters working in utilities, state energy offices, and/or utility commissions; independent system operators; regional transmission organizations; and regional energy and national energy planning organizations. The terminology used in demand forecasting and resource planning, baseline scenarios, historic and future planning time frames—as well as the format and definition and granularity of data requirements (e.g., coincident peak versus average peak demand reductions and source versus end-use energy savings)—can be different for the resource planner and forecaster than they are for a regulator of a retail energy provider.

7.9 DETERMINING NON-ENERGY BENEFITS

Virtually all energy efficiency programs have objectives associated with reducing energy use and costs. This guide focuses on documentation of these energy- and demand-related benefits and the associated avoided emissions. However, there is a wide range of other benefits that come from energy efficiency activities; for consumers, these non-energy benefits (NEBs) may actually drive their interest in efficiency investments. Historically, these NEBs have been, mostly, just subjectively noted as "other benefits," with little quantification or documentation of their value. Not putting a value to these NEBs, assuming they are positive, can result in negative bias in energy efficiency program investment decisions and less than fully effective program participation, designs, and marketing. The later would happen because the program implementers might not focus on the same benefits that participants focus on when they are making decisions about implementing efficiency projects.

NEBs can be categorized as those accruing to utilities (energy providers), society as a whole, and to individual participants.[80] Some research indicates that the value of benefits to society as a whole and individual participants make up the bulk of the value of NEBs.

NEW YORK STATE ENERGY RESEARCH AND DEVELOPMENT AUTHORITY EVALUATION OF NON-ENERGY BENEFITS

New York State Energy Research and Development Authority (NYSERDA) considers that non-energy benefits can be evaluated through a range of 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.
- Direct estimates (DE) of non-energy impacts can be made in conjunction with impact evaluation site visits, and generally focus on operations (e.g., water savings), materials, and labor hours due to reduced maintenance requirements. These estimates are based on equipment specifications, operating parameters, and other customer-supplied information, and they are monetized at the project level.

All of these approaches have benefits and drawbacks. The industry standard has been CV and DQ approaches. However, NYSERDA has pioneered the joint use of DQ and CA survey methods on its prior New York Energy $martSM Program and is mainly employing DE approaches currently.

DQ approaches have somewhat consistently produced NEB values around 50% of the value of the program-induced energy savings. CA approaches have resulted in similar results for some programs, but widely varied results for other programs ranging from 4% to 340% of the value of the program energy savings. DE approaches, which only measure the more quantifiable non-energy impacts, can result in relatively lower values as compared to DQ and CA approaches. NYSERDA's recent commercial/industrial NEBs research that found DE-based NEBs of $3 to $4 per natural gas MMBtu saved and $5 per MWh of electricity saved. The value of these NEBs to participants ranged from 6% to 39% of the retail value of their energy savings at typical New York utility rates.

Source: Jennifer Meissner, Program Manager, Evaluation, NYSERDA. 2012. References include: Energy & Resources Solutions, Inc. (ERS); Left Fork Energy; West Hill Energy & Computing; Megdal & Associates, LLC.; Opinion Dynamics Corporation. (June 2010). Report on Impact Assessments of Two NYSERDA Natural Gas Efficiency Programs: Con Edison Natural Gas Efficiency Program & National Grid Low Income Natural Gas Efficiency Program. Prepared for the New York State Energy Research and Development Authority (NYSERDA). http://documents.dps.ny.gov/public/Common/ViewDoc.aspx?DocRefId={CFCD7DE1-8B68-46CC-B64B-83ACD072A4FD}.

The following are examples of utility NEBs: [81]

- Avoided transmission and distribution capital and operating costs (particularly in areas with high energy use, high demand growth, and/or constrained distribution systems)
- Reduced line losses, voltage support (reliability), and power quality improvements
- Customer bill collection and service-related savings such as avoiding notices, shutoffs/reconnects, and carrying costs on arrearages.

The following are examples of societal NEBs:

- Economic development—for example, job development, both direct and indirect
- Energy price effects, such as stability, lower peak costs, and downward pressure on wholesale energy prices (although these price effects can also be considered "energy" benefits, these effects on marginal peak costs are not always included in benefit-cost analyses)
- Reduced air emissions, water use, water pollution, and the related emission trading values and/or health/hazard benefits
- National security improvements through reduced reliance on certain fuels.

Examples of efficiency program participant NEBs include the following:

- Indoor air quality improvements, improved comfort (e.g., quality of light, less noise, fewer drafts, better building temperature control), higher productivity and lower rates of absenteeism through better-performing energy using systems (e.g., ventilation, building shell, lighting)
- Reduced equipment operations and maintenance (O&M) costs because of more efficient, robust systems (although more complex systems could require more maintenance).

JOB IMPACTS OF CALIFORNIA ENERGY EFFICIENCY POLICIES

The following information is from a report presenting the results of the *California Workforce Education and Training Needs Assessment for Energy Efficiency, Demand Response, and Distributed Generation.* The assessment was conducted throughout calendar year 2010. Findings of the report are summarized in the table below and also include the following:

- By 2020, energy efficiency policies will result in about $11.2 billion of public and private investment, resulting in 211,000 direct, indirect, and induced jobs.
- Two-thirds of the direct jobs are in traditional construction trades, one-sixth are in building professional occupations.
- There are many more jobs are in traditional trades occupations than in new or specialized occupations.
- There are many more incumbent workers than new workers, which indicates the need for training in that there are relatively few slots for job seekers per year compared to the stock of current workers who may need skills upgrading.
- There are many more displaced construction workers than jobs, indicating that there are more unemployed, experienced construction workers due to the recession than there are job openings created by public/ratepayer investment in energy efficiency.

INDUSTRY/OCCUPATION GROUP	FULL-TIME EQUIVALENT JOB INCREASES	
	2015	2020
Residential Building Construction	5,072	7,104
Nonresidential Building Construction	5,342	6,924
Electrical Contractors	319	1,649
Plumbing, Heating, and AC Contractors	4,859	9,407
Drywall and Insulation Contractors	451	840
Manufacturing	51	574
Advertising and Related Services	956	1,794
Engineering Services, Architectural Services, etc.	2,118	4,026
Management, Public Administration	1,231	2,449
Office Administrative Services	2,021	3,958
All Other Industries	108	212
Total	22,528	38,937

The job increases are compared to the base year of 2009. They are in job person years (i.e., the number of full-time equivalent jobs for one year). They are not permanent jobs in that they depend on the annual investment of energy efficiency dollars.

Source: Zabin, C. (2011). California Workforce Education and Training Needs Assessment. Center for Labor Research & Education, Donald Vial Center on Employment in the Green Economy. www.irle.berkeley.edu/vial.

- Water and wastewater savings [82]
- Positive personal perceptions (e.g., "green," environmental consciousness) and for commercial businesses and public entities, improved public perceptions and the ability to market products and tenant leases
- Avoided capital cost for equipment or building component replacements whose capital costs can be paid from savings.

While most NEBs are considered to be positive, they can also be negative.[83] Examples of negative NEBs include aesthetic concerns associated with fluorescent lamps, "hassle" factor associated with implementing projects, and increased maintenance costs due to unfamiliarity with new energy efficient equipment or the need to operate more sophisticated equipment (e.g., building control systems) on a continuous basis.

As noted above, a common impact evaluation approach to NEBs is to list them as possible or probable benefits and not to quantify them. This is typically because of program administrators' (and sometimes evaluators') unfamiliarity with methods for quantifying these benefits, the cost of quantifying them, and the sense that the majority of economic benefits are associated with saved energy costs. However, the methods for documenting NEBs are improving and expertise in this area is increasing. And, perhaps most important, it is becoming increasingly clear that NEBs can have very high value for those making decisions about efficiency projects and to society as a whole.

In fact, for some programs, it appears that these participant NEBs can exceed the energy-related benefits.[84] For example, in a commercial office building where the cost of the employee salary and benefits per square foot dwarfs the cost of energy per square foot, an increase in employee productivity of a few percent may be significantly more valuable than decreasing energy costs by 30%. Including these non-energy benefits in evaluations can be quite valuable, given that energy efficiency programs increasingly are emphasizing these types of participant non-energy benefits when marketing energy efficiency programs to customers. Those factors arguably play an important role in persuading customers to make the significant investments necessary to achieve comprehensive energy savings in a home or business.

In terms of societal NEBs, one benefit that has generated a great deal of interest is job creation. Jobs are created as a result of efficiency programs in three categories (see the California Efficiency Policy sidebar for examples of results of jobs analyses):

- **Direct.** Direct jobs are in firms that are actually receiving the efficiency program dollars and doing the energy efficiency work that a program is targeting (e.g., construction, engineering, architecture).
- **Indirect.** Indirect jobs are jobs in firms supplying goods and services to energy efficiency firms (e.g., manufacturing, accounting).
- **Induced.** Induced jobs are those created by the demand generated by wage and business income from energy efficiency investments and by energy bill savings (e.g., jobs induced by NEBs, such as grocery store or apparel).

The methods for documenting NEBs tend to fall into one or more of the following three categories:

- **Measurement of benefits.** These methods are used with benefits that can be directed, measured, or calculated, such as water savings. As with energy savings, a variety of "measurement" approaches are possible, including the use of control groups and M&V. In some cases, the measurements can be made of secondary metrics that are indicators. For example, comfort can be defined via monitoring of indoor temperatures and humidity to document whether participants' homes, compared with those of non-participants, are within defined comfort conditions.[85]
- **Modeling.** These methods include macroeconomic modeling and analysis tools that look at broader societal impacts such as job growth or modeled estimates of emissions impacts.
- **Surveys.** These are used for documenting many different types of NEBs. There is a wide range of surveying approaches to determine these benefits. These include willingness to pay and willingness to accept contingent valuation (CV) studies, comparative or relative valuations, and other revealed preference and stated preference approaches. Surveys are used specifically for determining relatively subjective program participant benefits when quantification is difficult and/or expensive.[86] However, surveys can be used for almost all benefit types where participants and non-participants can be asked to provide data (e.g., how many people they hired for determining job impacts, whether they believe their indoor air quality is better, if there are distribution projects that were delayed).

Currently, several states are including NEBs in their evaluations, but not many. In particular for cost-effectiveness analyses, the ACEEE 2012 review of evaluation practices[87] indicated the following:

> …. while 36 states (including all the states with TRC [total resource cost] as their primary [cost-effectiveness] test) treated "participant costs" for the energy efficiency measures as a cost, only 12 states treated any type of participant "non-energy

> benefits" as a benefit. … most of those "non-energy" participant benefits were confined to "water and other fuel savings." Only 2 states quantified a benefit for "participant O&M savings" and none quantified any benefits for things like "comfort," "heath," "safety," or "improved productivity" in their primary benefit-cost test.

Given the potential significant value of NEBs, it is expected that more jurisdictions will analyze NEBs and that more cost-effectiveness analyses will take them into consideration, such as in the societal cost test.

7.10 IMPACT EVALUATION FOR UNIQUE ENERGY EFFICIENCY PROGRAM TYPES

7.10.1 Residential Behavior-Based Programs

Encouraging people and organizations to do things differently is a tricky proposition. Energy efficiency is not just about new technologies; it is about new behaviors and better decisions. For more than three decades, program administrators have offered consumers energy efficiency programs that have used strategies such as subsidies, rebates, or other financial incentives to motivate consumers to install technologies and high-efficiency measures. In many respects, these programs have focused on affecting behaviors by giving people incentives to undertake efficiency actions.

However, in the last several years, there has been substantial interest in broadening energy efficiency program portfolios (particularly the residential programs) to include behavior-based programs that use strategies intended to affect how consumers use energy in order to achieve energy and/or demand savings. These programs typically include outreach, education, competition, rewards, benchmarking, and/or feedback elements. In some cases, this new generation of programs takes advantage of technological advances in Internet and wireless communication to find innovative ways to both capture energy data at a higher temporal and spatial resolution than ever before and to communicate the energy data to households in creative new ways that leverage social science-based motivational techniques.

These programs are unique in that they may rely on changes to consumers' habitual behaviors (e.g., turning off lights) or "one-time" behaviors (e.g., changing thermostat settings). In addition, these programs may target purchasing behaviors (e.g., purchases of energy-efficient products or services), often in combination with other programs such as rebate programs or direct install programs. These programs are also distinguished by normally being evaluated using large-scale data analysis approaches involving randomized controlled trials or quasi-experimental methods versus deemed savings or M&V approaches.

Obstacles to the widespread adoption of behavior-based programs include issues relating to whether these programs can be evaluated in a rigorous way, the savings persist, and the evaluated results shown for one program can be applied to another program. Another SEE Action report[88] specifically addresses behavior-based programs and prepares recommendations with respect to these issues. The following lists the primary recommendations from that report:

- For program evaluation design, the use of randomized controlled trials (RCT) is recommended. These will result in robust, unbiased estimates of program energy savings. If this is not feasible, it is suggested that "quasi-experimental" approaches be used.
- For a level of precision that is considered acceptable in behavioral sciences research, it is recommended that a null hypothesis (e.g., a required threshold such as the percent savings needed for the benefits of the program to be considered cost-effective) should be established. The program savings estimate should be considered acceptable (i.e., the null hypothesis should be rejected) if the estimate is statistically significant at the 5% level or lower.
- In order to avoid potential evaluator conflicts of interest, it is recommended that results are reported to all interested parties and that an independent third-party evaluator transparently defines and implements the following:
  - Analysis and evaluation of program impacts
  - Assignment of facilities (e.g., households) to treatment and control groups (whether randomly assigned or matched)
  - Selection of raw utility data to use in the analysis
  - Identification and treatment of missing values and outliers
  - Normalization of billing cycle days
  - Identification and treatment of households that close their accounts.
- For the analyses of savings, it is recommended to use a panel data model[89] that compares the change in energy use for the treatment group to the change in energy use for the control group, especially if the evaluation design is quasi-experimental.

With respect to external validity of behavior-based program evaluation results, it is possible in theory that a predictive model could be created that allows program estimates to be extrapolated to future years and new populations without actually measuring the savings estimates in those years. It is also possible that behavior-based programs could move to a deemed or modeled savings approach over time. However, it is generally believed that the industry is not yet at this point, due to the small number of behavior-based programs that have been evaluated using rigorous approaches.

For more details and recommendations on evaluating behavior-based programs, please see the 2012 SEE Action report referenced above. Another source of information on behavior-based programs and their evaluation are the proceedings of the Behavior, Energy, and Climate Change Conferences (http://beccconference.org/).

7.10.2 Education and Training (E&T) Programs

Education and training (E&T) programs are seen as very important strategies for expanding energy efficiency's reach as a sustained, long-term resource. Education and training programs may be targeted to either end-use customers or other market actors (e.g., trade allies) whose activities influence the energy-related choices of end-use customers. These programs can include advertising, public service announcements, education efforts, training activities (e.g., for contractors, building operators, and designers), outreach efforts, demonstration projects, and other information or communication-based efforts.

Typically, E&T programs have one or more of the following general goals or desired outcomes:

- 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.
- Train contractors, engineers, architects, and building operators on skills, best practices, tools, and other issues related to energy efficiency. An example program is the Building Operator Certification Program (www.theboc.info).

Almost every energy efficiency program provides some level of educational and/or informational content. However, education- or training-specific programs are typically designed to achieve energy or demand savings indirectly. Thus, while they are important strategies, these programs only indirectly result in energy and demand savings and therefore represent unique impact evaluation challenges.

For E&T programs, evaluations usually focus on documenting the degree to which the programs are achieving their desired outcomes 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 previously cited California Energy Efficiency Evaluation 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' 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 (education) evaluation community.

7.10.3 Market Transformation (MT) Programs

There are many definitions of market transformation (MT), although it is often considered the ultimate goal of publicly and consumer-funded efficiency programs. In this guide, the definition of *market transformation* is: "a reduction in market barriers resulting from a market intervention, as evidenced by a set of market effects (or perhaps more specifically a set of market progress indicators) that are likely to last after the intervention has been withdrawn, reduced, or changed." MT denotes a permanent, or at least long-term, change in the operation of the market for energy efficiency products and services. As such, their "end-point" can be considered to have occurred if the subject energy efficiency measure(s) or practices are either simply part of common practice (i.e., all consumers, designers, builders, and operators just "do it") or required per a code or standard.

Market transformation programs attempt to reduce market barriers through market interventions. During the 1990s, the focus of many energy efficiency efforts shifted from resource acquisition to market transformation. Subsequently, there was a shift back to more resource acquisition-focused programs that did not necessarily include market intervention components, or at least did not include components defined as such. However, current best practices have all publicly or energy consumer-funded programs having at least some MT elements, in that they involve changing how energy efficiency activities take place in the marketplace. Thus, MT and other program types are now often implemented in a complementary manner.

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. 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. Also, MT programs are dynamic, and thus the nature of market effects can be expected to vary over time. Thus, market effects evaluations, conducted at several points in time, are the primary evaluation activity usually associated with MT programs. See Appendix B for a description of these types of evaluations.

Evaluation of MT interventions also needs to focus on the mechanisms through which changes in adoptions and energy use are ultimately induced. This means that considerable attention must be focused on indicators of market effects through market tracking. For example, an MT evaluation might first report changes in sales patterns and volumes for particular energy efficiency products as an indication of program progress in meeting program goals.

Evaluation also 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, because the interest is in long-term changes in the market versus more immediate and direct energy savings for resource acquisition programs.

As a final note for MT program evaluations, and as discussed briefly in Appendix B and the sidebar "Theory-Based Evaluation: A Guiding Principle for Market Transformation Evaluation," a program logic model and a program theory are common components of MT programs and their evaluations. A program logic model is a visual representation of the program's theory that illustrates a set of interrelated program activities that combine to produce a variety of outputs that lead to key (in these programs) MT outcomes. Thus, MT program evaluations should entail collecting information that can be used to refine the underlying program theory (see sidebar).

7.10.4 Codes and Standards Programs

As mentioned in the previous section, from a market transformation perspective, the end goal of energy efficiency programs is to have effective energy efficiency actions become either common practice or established in energy codes or standards (C&S). Building energy codes and standards set minimum requirements for energy efficient design and construction for new and renovated buildings. Appliance (equipment) standards set minimum requirements for energy consumption of appliances and equipment. States and local governments generally establish building energy codes applicable in their jurisdictions and both states (approximately 15) and the federal government establish equipment standards.[90]

Local governments, states, and the federal government also establish energy efficiency programs to support the development of, implementation of, and compliance with C&S. These programs include efforts such as emerging technology programs, compliance-enhancement programs, and stretch (or reach) goal programs, as well as training on C&S for building code officials, builders, contractors, and designers. States are interested in C&S and efforts/programs to support them such as increased code enforcement, because these support energy savings goals as well as air pollution reduction and GHG mitigation goals.

When determining energy (and demand) savings from C&S, there are two aspects to the project scenarios: C&S requirements and C&S enforcement/compliance (noting that enforcement is not the same as compliance, compliance being the actual goal that results in energy savings). Whether a C&S is actually being complied with can be determined through surveys of practices in the field. This can be done through review of permits. However, that can be quite difficult, given the lack of digitized and organized permits in many jurisdictions and because permits are often not "pulled" for many projects such as small HVAC and lighting retrofits. Also, "percent compliance" is not a very definable concept, given that compliance is usually not a "yes or no" issue, as partial compliance is often the case. Thus, energy performance metrics, such as energy use per square foot for the buildings of interest (over an extended period of time), can be a much better targeted metric for gauging whether C&S are being complied with and to see if there are any energy savings taking place.

The approaches to determining energy savings associated with C&S efforts can involve deemed values, M&V approaches, and large-scale consumption data analyses approaches. However, the usual approach is a combination of calculations and field data collection that involves the following:

- Establishing an efficiency baseline for each energy efficiency measure based on analyses of common practices or naturally occurring market adoption rate (NOMAD)
- Defining unit energy savings per measure per year based on C&S compliant measures' versus baseline measures' efficiency/performance
- Defining a market volume baseline for how many units per year are installed in the marketplace
- Calculating an energy and demand savings value
- Adjusting for compliance rates (which for a specific code or standard may be the compliance rate found in the market or, for a program that supports C&S, may be the improved rate as a result of the program, compared against a baseline compliance rate).

Figure 7.2 illustrates the components of C&S program evaluation.

In terms of programs intended to improve C&S compliance, the ultimate measure for such efforts is of course better compliance. However, intermediate metrics may also be important to analyze, particularly given the long-term perspective associated with a C&S effort. The following questions can be raised to discover more about these intermediate metrics:

- Are training programs being offered? What are the results of pre- and post-test knowledge changes as a result of the training? How many trained builders and building officials are there as a result of the program?
- Have there been reach/stretch C&S established, and what is the participation level as a result of the program?
- Are there changes in attitudes in code enforcement offices and/or perceptions of improved enforcement efforts from code officials?

When attempting to attribute achieved savings from C&S to a specific program or group of programs, it is important to review the entire C&S process. This can include interviewing stakeholders, assessing their effectiveness, and documenting the program contributions. In these attribution efforts, evaluators often use a preponderance-of-evidence approach to forming a judgment, recognizing that attribution is elusive when so many parties are involved.

THEORY-BASED EVALUATION: A GUIDING PRINCIPLE FOR MARKET TRANSFORMATION EVALUATION

Theory-based evaluation (TBE), an evaluation approach that has been widely used in the evaluation of social programs in other fields, has seen increased use in the energy-efficiency industry. 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 congruence 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 measure installation, resulting in 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.

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, involves a relatively large number of causal steps and mechanisms, and encompasses changing the behavior of multiple categories of market actors—all of which makes it particularly important to focus on specifying and testing a detailed and articulated program theory.

7.10.5 Demand Response (DR) Programs

Demand response (DR) is the reduction of a consumer's energy use at times of peak use in order to help system reliability, reflect market conditions and pricing, or support infrastructure optimization or deferral of additional infrastructure. DR programs may include contractually obligated or voluntary curtailment, direct load control, and pricing strategies. As noted in the demand savings section above, DR programs reduce a utility customer's electricity demand in response to dispatch instructions or price signals sent to the program participants—a "call" for reductions. Thus, the DR peak demand savings is determined for when there is a "call" for program participants to reduce their energy consumption rate. When this call occurs is referred to as the DR "event."

Demand response includes an array of programs relying on various incentive mechanisms. Some DR programs simply involve a contract with a utility customer to reduce their load when asked (or told) to do so; this is historically called *interruptible electricity contracts* or *rates.* Other DR programs include residential air-conditioner cycling, where demand savings result from a load control device that actively reduces electricity demand when activated by a command from a central control center. Dynamic pricing is another form of DR, where a price signal is sent to consumers, with the expectation that they will respond to a higher price by reducing consumption (and therefore demand). Most DR programs require active and continued participation because it is assumed that once a participant is no longer enrolled in the program, electric loads revert back to normal patterns.

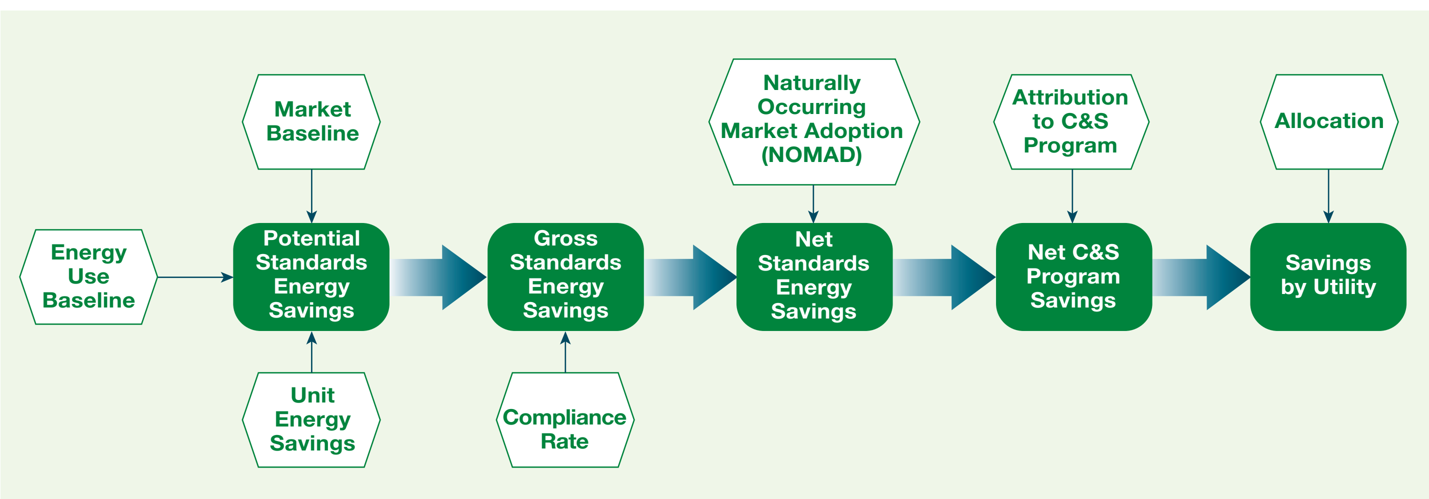
In most cases, DR programs incorporate and rely on electric consumption data for the participants being collected and recorded on at least a 15-minute basis. Therefore, demand is continuously monitored, including during the DR events. The savings are the difference between the electricity consumption (demand) during an event and the baseline.

Thus, the pertinent question about DR program evaluations is a common one: "What is the baseline?" For DR programs, the baseline is defined as part of the program design, and because participants are typically paid on performance (i.e., how much they reduce their demand), the baseline is also defined in the DR participant contract.[91] Typical baselines are the demand for energy the hour before the event, the hour after an event, the average of hour before and after, or the demand for a day similar to the one when the event occurred (e.g., if the event was on a summer Tuesday afternoon, the baseline might be the average of all Tuesday afternoons that summer). In some cases, the baseline might be determined with a control group, such as if a large-scale consumption data evaluation approach is used and the metric of interest is demand reduction for all of the participants, not just individual participants (this might be the case for evaluation of a rate schedule used to induce DR in consumers).

7.10.6 Greenhouse Gas (GHG) Mitigation Programs

In Chapter 6, the documenting of avoided emissions—including greenhouse gases (GHG)—that result from energy efficiency programs is described, including issues associated with documenting avoided emissions in cap-and-trade programs and determining

FIGURE 7.2: Components of C&S savings estimate



Source: California Public Utilities Commission (CPUC). (April 9, 2010). Codes & Standards (C&S) Programs Impact Evaluation: California Investor Owned Utilities' Codes and Standards Program Evaluation for Program Years 2006–2008. CALMAC Study ID: CPU0030.06 Final Evaluation Report. Prepared by KEMA, Inc.; The Cadmus Group, Inc.; Itron, Inc.; Nexus Market Research, Inc.; and ENRG, Inc. for the CPUC Energy Division. www.calmac.org/publications/Codes_Standards_Vol_III_FinalEvaluationReportUpdated_04122010.pdf.

whether the avoided emissions data are just for informational purposes or are part of a benefit-cost analysis or regulatory requirement. While avoided emissions can be part of the benefits of an energy efficiency program, end-use energy efficiency activities can also be part of a GHG mitigation program.

End-use energy efficiency represents a significant, cost-effective approach to reducing GHG emissions. This has been well documented in studies by the Intergovernmental Panel on Climate Change and numerous other government, private, and non-governmental organizations.[92] However, as with the implementation of energy efficiency programs solely for the purpose of energy use and demand reductions, energy efficiency for GHG mitigation suffers from the difficulty of cost-effectively documenting the benefits of energy efficiency activities at levels of certainty that can satisfy regulators, investors, and project proponents.

Greenhouse gas voluntary and mandatory mitigation programs in the United States and internationally include energy efficiency as a mitigation strategy. Examples include the Verified Carbon Standard (VCS, http://v-c-s.org), the Gold Standard (www.cdmgoldstandard.org), the Regional Greenhouse Gas Initiative (RGGI, www.rggi.org), and the Clean Development Mechanism (CDM, http://cdm.unfccc.int/index.html). In these programs, end-use efficiency is either an eligible activity for generating emissions offsets (as in VCS, Gold Standard, and CDM) or a strategy for investing funds generated through the program (as with the use of allowance auction funds in RGGI).

GHG Offsets

The CDM is the world's largest GHG offset program. It is a mechanism under the Kyoto Protocol, which set targets for industrialized countries to reduce their domestic emissions as part of the United Nations Framework Convention on Climate Change. The basic concept of the CDM is that emissions offsets from projects in developing countries without a commitment to reduce emissions are sold to entities in developed countries with emissions-reduction commitments at a cost less than these entities' costs to reduce their own emissions. The money generated from the sale of offsets pays for projects in the developing countries. Greenhouse gas emissions are reduced in an economically efficient manner, and developing countries receive funds to invest in sustainable development projects.[93]

As part of the CDM implementation, methodologies (per the terminology of this guide, project-specific M&V plans) have been created for documenting GHG reductions from a wide range of energy efficiency activities. These cover commercial, industrial, and municipal facilities (retrofits and new construction projects associated with steam systems, water pumping, energy recovery, boilers, chillers, street lighting, district heating, metals production, and agriculture) and households (retrofits and new construction projects associated with cook stoves, water pumping, water purification, refrigerators, lighting, and whole-building projects). These methodologies and other tools for documenting GHG reductions from energy efficiency projects are available at the CDM website.

Unlike the CDM, which is regulated and part of a mandatory GHG emissions program, the VCS is a greenhouse gas accounting program used by projects around the world to verify and issue carbon credits in voluntary markets. It is one of several voluntary GHG accounting programs in place throughout the world. Another is the Gold Standard, which is a carbon-offset standard that certifies compliance credits created through the CDM and voluntary carbon credit programs. The VCS and the Gold Standard also have methodologies that can be used to document emission reductions from energy efficiency activities. These methodologies, as well as the CDM methodologies, can be useful tools for evaluators of energy efficiency programs seeking to document GHG and other emission reductions.

Investing GHG Funds in Energy Efficiency

There are at least two ways to think about the relationship between end-use efficiency and GHG reductions: (1) assume that energy efficiency is an important societal goal and ask whether GHG reduction programs can lend value to its attainment; or (2) focus on GHG reduction as the goal and ask whether accelerated energy efficiency is a central element in attaining those essential reductions. While both goals—efficient use of energy and climate mitigation—are important, in the second context, even though energy efficiency is essential to GHG attainment, it does not follow that simply monetizing GHG emissions will actually call forth the full level of energy efficiency that is both attainable and cost-effective, given the variety of market barriers associated with energy efficiency.[94]

However, auctioning cap-and-trade allowances can raise, in aggregate, large amounts of funds that can be used in complementary programs that do address these barriers. [95] These funds can be used in a wide range of efficiency support activities such as public benefit programs; energy efficiency resource standards; appliance efficiency standards; building energy codes; and research, development, and demonstration.

One example of this approach is what was done the Regional Greenhouse Gas Initiative (RGGI), which was the first market-based regulatory program in the United States to reduce greenhouse gas emissions. It is a cooperative effort among the states of Connecticut, Delaware, Maine, Maryland, Massachusetts, New Hampshire, New York, Rhode Island, and Vermont. Together, these states have capped

REGIONAL GREENHOUSE GAS INITIATIVE

In 2009, ten Northeastern and Mid-Atlantic states began the Regional Greenhouse Gas Initiative (RGGI), the country's first market-based program to reduce emissions of carbon dioxide (CO_2) from power plants. There are now nine states in the program.

RGGI Produced New Jobs

Taking into account consumer gains, lower producer revenues, and net positive macroeconomic impacts, RGGI led to overall job increases amounting to thousands of new jobs over time. RGGI job impacts may in some cases be permanent; others may be part-time or temporary. But according to our analysis, the net effect is that the first three years of RGGI led to over 16,000 new "job years," with each of the original ten states showing net job additions. Jobs related to RGGI activities are located around the economy, with examples including engineers who perform efficiency audits; workers who install energy efficiency measures in commercial buildings; staff performing teacher training on energy issues; or workers in state-funded programs that might have been cut had a state not used RGGI funds to close budget gaps.

Energy Efficiency in RGGI

The states have used their RGGI dollars very differently, in ways that affect the net benefits within the electric sector and in the larger state economy. While all states originally committed to using at least 25 % of auction proceeds for "public benefit or strategic energy" purposes, some states contributed a much larger amount to those ends. But from a strictly economic perspective, some uses of proceeds clearly deliver economic returns more readily and substantially than others. For example, RGGI-funded expenditures on energy efficiency depress regional electrical demand, power prices, and consumer payments for electricity. This benefits all consumers through downward pressure on wholesale prices, even as it particularly benefits those consumers that actually take advantage of such programs, implement energy efficiency measures, and lower both their overall energy use and monthly energy bills. These savings stay in the pockets of electricity users directly. But there are also positive macroeconomic impacts as well: the lower energy costs flow through the economy as collateral reductions in natural gas and oil in buildings and increased consumer disposable income (from fewer dollars spent on energy bills), lower payments to out-of-state energy suppliers, and increased local spending or savings. Consequently, there are multiple ways that investments in energy efficiency lead to positive economic impacts; this reinvestment thus stands out as the most economically beneficial use of RGGI dollars.

Source: Hibbard, P.; Tierney, S.; Okie, A.; Darling, P. (November 2011). The Economic Impacts of the Regional Greenhouse Gas Initiative on Ten Northeast and Mid-Atlantic States, Review of the Use of RGGI Auction Proceeds from the First Three-Year Compliance Period. The Analysis Group. www.analysisgroup.com/uploadedfiles/publishing/articles/economic_impact_rggi_report.pdf.

and will reduce CO_2 emissions from the power sector 10% by 2018. States sell nearly all emissions allowances through auctions and invest proceeds in consumer benefits such as energy efficiency, renewable energy, and other clean energy technologies.

The sidebar "Regional Greenhouse Gas Initiative" summarizes the results of an impact evaluation of the benefits of the RGGI investments in energy efficiency. As of the writing of this guide, the State of California is also reviewing whether to use its new cap-and-trade GHG program allowance proceeds in a similar manner to the RGGI. These programs and their evaluations represent new opportunities for energy efficiency and energy efficiency evaluations.

Chapter 7: Notes

[68] Electricity system operators are Independent System Operators (ISO) or Regional Transmission Organizations (RTO) that operate a regional energy market, capacity market, or both. A summary document on the subject of energy efficiency and forward capacity markets is Gottstein, M.; Schwartz, L. (2010). *The Role of Forward Capacity Markets in Increasing Demand-Side and Other Low-Carbon Resources Experience and Prospects.* Regulatory Assistance Project (RAP). www.raponline.org/docs/RAP_Gottstein_Schwartz_RoleofFCM_ExperienceandProspects2_2010_05_04.pdf.

[69] DEER contains information on selected energy-efficient technologies and measures. It provides estimates of the energy-savings potential for these technologies in residential and nonresidential applications. The database contains information on typical measures and data on the costs and benefits of more energy-efficient measures. www.deeresources.com.

[70] Vine, E.; Hall, N.; Keating, K.M.; Kushler, M.; Prahl, R. (August 2011). "Emerging Evaluation Issues Revisited." *2011 International Energy Program Evaluation Conference Proceedings,* August 16-18, 2011, Boston, Massachusetts. Madison, Wisconsin: International Energy Program Evaluation Conference (IEPEC). www.iepec.org/2011PapersTOC/papers/096.pdf#page=1.

[71] Skumatz, L.A.; Khawaya, M.; Colby, J. (2009). *Lessons Learned and Next Steps in Energy Efficiency Measurement and Attribution: Energy Savings, Net to Gross, Non-Energy Benefits, and Persistence of Energy Efficiency Behavior.* Prepared for the California Institute for Energy and Environment (CIEE) Behavior and Energy Program. http://uc-ciee.org/downloads/EEM_A.pdf.

[72] One such study is Hirst, E.; White, D. (1985). *Indoor Temperature Changes After Retrofit: Inferences Based on Electricity Billing Data for Nonparticipants and Participants in the BPA Residential Weatherization Program.* Oak Ridge National Laboratory. www.ornl.gov/info/reports/1985/3445600444954.pdf.

[73] California Public Utilities Commission (CPUC). (April 2006). *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals.* www.calmac.org/events/EvaluatorsProtocols_Final_AdoptedviaRuling_06-19-2006.pdf.

[74] "Uniform Methods Project." (2012). U.S. Department of Energy. www1.eere.energy.gov/deployment/ump.html. Includes information on steps to increase the accuracy of evaluation results, particularly with respect to sampling.

[75] Note the counterintuitive implication of this standard definition. Low precision values correspond to narrow intervals and, hence, describe tight estimates. This can lead to confusion when estimates are described as having "low precision."

[76] Although there is a close relationship between confidence and precision, they are not direct complements of each other. If the confidence level is 90%, there is no reason that the precision needs to be 10%. It is just as logical to talk about 90/20 confidence and precision as 90/10.

[77] National Action Plan for Energy Efficiency. (2007). *Guide to Resource Planning with Energy Efficiency.* Prepared by Snuller Price et al., Energy and Environmental Economics, Inc. (E3). www.epa.gov/cleanenergy/documents/suca/resource_planning.pdf. A companion document to the original version of this guide, which addresses resource planning and energy efficiency.

[78] American Council for an Energy-Efficient Economy (ACEEE). (October 2011). *State Energy Efficiency Resource Standard (EERS) Activity.* http://aceee.org/files/pdf/policy-brief/State EERS Summary October 2011.pdf. EERS and other energy efficiency policies are described in Section 2.2.1. An up-to-date list of states that currently maintain an EERS can also be found in the Database of State Incentives for Renewable Energy (DSIRE) at www.dsireusa.org.

[79] Barbose, G.; Billingsley; M.; Goldman, C.; Hoffman, I.; Schlegel, J. (August 2012). "On a Rising Tide: The Future of U.S. Utility Customer-Funded Energy Efficiency Programs." *2012 ACEEE Summer Study Proceedings;* August 12-17, 2012, Pacific Grove, California. Washington, D.C.: American Council for an Energy-Efficient Economy (ACEEE). LBNL-5755E. www.aceee.org/files/proceedings/2012/data/papers/0193-000173.pdf.

[80] Categories based on: Skumatz, L.A. (1997). "Recognizing All Program Benefits: Estimating the Non-Energy Benefits of PG&E's Venture Partners Pilot Program (VPP)." *Proceedings of the 1997 Energy Evaluation Conference.* August 27-29, 1997, Chicago, Illinois. Madison, Wisconsin: International Energy Program Evaluation Conference (IEPEC). www.iepec.org/1997PapersTOC/papers/033.pdf.

[81] Skumatz, L.A.; Khawaja, M.S.; Krop, R. (May 2010). *Non-Energy Benefits: Status, Findings, Next Steps, and Implications for Low Income Program Analyses in California.* Revised Report. Prepared for Brenda Gettig, Sempra Utilities. www.liob.org/docs/LIEE Non-Energy Benefits Revised Report.pdf.

[82] Water savings can come from energy efficiency measures, such as low-flow showerheads and cooling tower upgrades that are designed to save energy but also reduce water (and wastewater flows). In addition, there are water-saving projects (reducing irrigation needs, process water requirements) that save energy via reduced water pumping, conditioning, and/or heating.

[83] In fact, in a strict sense, when used in association with incremental savings for programs or measures, NEBs should be described as "net" of what would have occurred absent the energy efficiency programs; thus, in some cases they might be positive in an absolute sense, but negative compared to, for example, what would have occurred if the consumer had purchased a standard energy efficiency product.

[84] Skumatz, L.A.; Khawaja, M.S.; Krop, R. (May 2010). *Non-Energy Benefits: Status, Findings, Next Steps, and Implications for Low Income Program Analyses in California.* Revised Report. Prepared for Brenda Gettig, Sempra Utilities. www.liob.org/docs/LIEE Non-Energy Benefits Revised Report.pdf.

[85] As an example of a quantifiable definition of "comfort," ANSI/ ASHRAE Standard 55-2010 defines a range of indoor thermal environmental conditions acceptable to a majority of occupants.

[86] Or when the "perception" is the factor that affects the benefit (such as comfort), or the associated perception of a return on investment that may affect program "uptake."

[87] Kushler, M.; Nowak, S.; Witte, P. (February 2012). *A National Survey of State Policies and Practices for the Evaluation of Ratepayer-Funded Energy Efficiency Programs.* American Council for an Energy-Efficient Economy (ACEEE). Report Number U122. www.aceee.org/research-report/u122.

[88] State and Local Energy Efficiency Action Network. (May 2012). *Evaluation, Measurement, and Verification (EM&V) of Residential Behavior-Based Energy Efficiency Programs: Issues and Recommendations.* Prepared by Todd, A.; Stuart, E.; Schiller, S.; Goldman, C.; Lawrence Berkeley National Laboratory. www1.eere.energy.gov/seeaction/pdfs/emv_behaviorbased_eeprograms.pdf.

[89] A panel data model is an analysis model in which many data points over time are observed for a certain population (also called a time-series of cross-sections).

[90] "Building Energy Codes Program." (2012). U.S. Department of Energy. www.energycodes.gov.

[91] Some DR programs might actually be set up with participants receiving a discounted cost of energy throughout the year, and then penalties if they do not reduce their demand when called upon to do so.

[92] Intergovernmental Panel on Climate Change (IPCC). (2007). *Pathways to a Low-Carbon Economy: Version 2 of the Global Greenhouse Gas Abatement Cost Curve*. Prepared by McKinsey & Company. www.epa.gov/statelocalclimate/documents/pdf/mckinsey_summary_11-19-09.pdf.

Intergovernmental Panel on Climate Change (IPCC). (2007). *Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, 2007. www.ipcc.ch/publications_and_data/ar4/wg3/en/contents.html

[93] Schiller, S.R. (August 2011). "Documenting Greenhouse Gas Emission Reductions from End-Use Efficiency Activities in the Kyoto Protocol's Clean Development Mechanism Offset Program." *2011 International Energy Program Evaluation Conference Proceedings,* August 16-18, 2011, Boston, Massachusetts. Madison, Wisconsin: International Energy Program Evaluation Conference (IEPEC). www.iepec.org/2011PapersTOC/papers/115.pdf#page=1.

[94] Schiller, S.R.; Prindle, B.; Cowart, R.; Rosenfeld, A.H. (2008). "Energy Efficiency and Climate Change Mitigation Policy." *2008 ACEEE Summer Study Proceedings.* August 17-22, 2008, Pacific Grove, California. Washington, DC: American Council for an Energy-Efficient Economy (ACEEE). www.aceee.org/files/proceedings/2008/data/papers/8_306.pdf.

[95] Under a cap-and-trade program, an overall emission tonnage cap is set for an affected sector or set of emission sources. Governments create allowances representing the temporary right to emit one unit (e.g., one ton) within the total cap amount. Initial allowance allocations can be sold (auctioned) by the government or distributed for free to affected sources or to others. In the above-referenced paper (Schiller et. al.), the authors call the activities paid for by the allowance money *complementary policies and programs*—in other words, programs that while running parallel to cap-and-trade programs are still integral to the success of cap-and-trade and GHG mitigation policies.

Chapter 8

Impact Evaluation Planning

Chapter 8 builds on preceding chapters and presents impact evaluation concepts and the steps involved in the planning process. These include the development of evaluation approaches, budgets, and schedules. The first section discusses how evaluation planning and reporting is integrated into the program implementation process, while the second section presents the concept of a hierarchy of planning documents, starting with an evaluation framework. The third section presents 14 issues and questions that help determine the scope and scale of an impact evaluation. The last section provides some guidance on preparing program impact evaluation plans and site-specific measurement and verification (M&V) plans, as well as a checklist for preparing an evaluation plan.

8.1 INTEGRATION OF EVALUATION INTO THE PORTFOLIO IMPLEMENTATION CYCLE

Before describing the evaluation planning process, it is important to understand how it is integral to what is typically a cyclic program or portfolio planning-implementation-evaluation process. In most cases, the overall cycle time frame is determined by a program's—or more likely a portfolio's—funding and contracting schedules (cycles). These portfolio implementation cycles can be one or two years or even longer. The point at which programs are being designed is ideally when the impact, market, and process evaluation planning process should begin. This is primarily so that the program budget, schedule, and resources can properly take into account evaluation requirements and opportunities. It is also a way to reinforce the concept that evaluation is an integral part of the portfolio process, supporting the portfolio's success through an assessment of the program's impacts as well as the program's theory for how savings are to be achieved.[96] *In addition, early consideration of evaluation, on a very practical level, helps ensure that the data collection required to support expected evaluation efforts is accommodated at the time of implementation.* It is also helpful to decide early in the process which entities will collect and analyze data (e.g., program implementers, an independent third-party evaluator, or both).

Evaluations should be produced within a portfolio cycle or very soon after the completion of a cycle. This is so evaluation results can document the operations and effects of the programs in a timely manner and provide feedback for ongoing program improvement, provide information to support energy efficiency portfolio assessments (including market assessments and potential studies), and help support the planning of future portfolio cycles, load forecasts, and energy resource plans. For impact evaluations that examine energy savings of certain measures and program mechanisms, the evaluation information can also be used to inform deemed savings values though updating of technical reference manuals (TRMs).

Figure 8.1 shows the energy efficiency program implementation cycle, emphasizing evaluation activities and feedback to the current and future programs (portfolio). This figure and the following discussion concerning evaluation timing are related to the activities associated with formal impact evaluations, usually conducted by an independent third-party evaluator. There are other documentation activities that are carried out by program implementers. These include defining baselines, conducting project inspections, and calculating savings for the paying of incentives to contractors and preparing their own (implementer) estimates of claimed savings. There is an interaction between these different activities by the implementer and evaluator, which is discussed in Section 8.3; however, this text is primarily focused on the activities carried out by just the evaluator.

The first three evaluation activity steps displayed in Figure 8.1 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. For example, if the program goal is to save electricity during peak-use periods, the evaluation goals can include accurately documenting how much electricity is saved during a defined peak period (gross impact) and how much of these savings can be attributed to the program (net impact). High-level evaluation schedules and budgets can also be established at this time.
- **Program design.** The evaluation design effort should begin during the program design phase. The objective should be a preliminary evaluation plan that includes a schedule and a budget as well as a description of each work task to be completed. The evaluation issues described in Section 8.3 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 necessary priority given to the program design and the need to have the program plan pretty well

completed before the evaluation plan can be finalized. Another reason why the evaluation design may very well not be completed at this time is that it is not unusual, although not best practice, to select the evaluator after the program has been fully designed (and sometimes after it is implemented). Nonetheless, it is ideal to assign evaluation staff or consultants to the program evaluation during program design activities. Regardless of when the formal evaluator gets involved, evaluation goals and metrics should be considered, and an initial set of evaluation priorities (with a rough schedule and budget) should be established during this portfolio cycle phase.

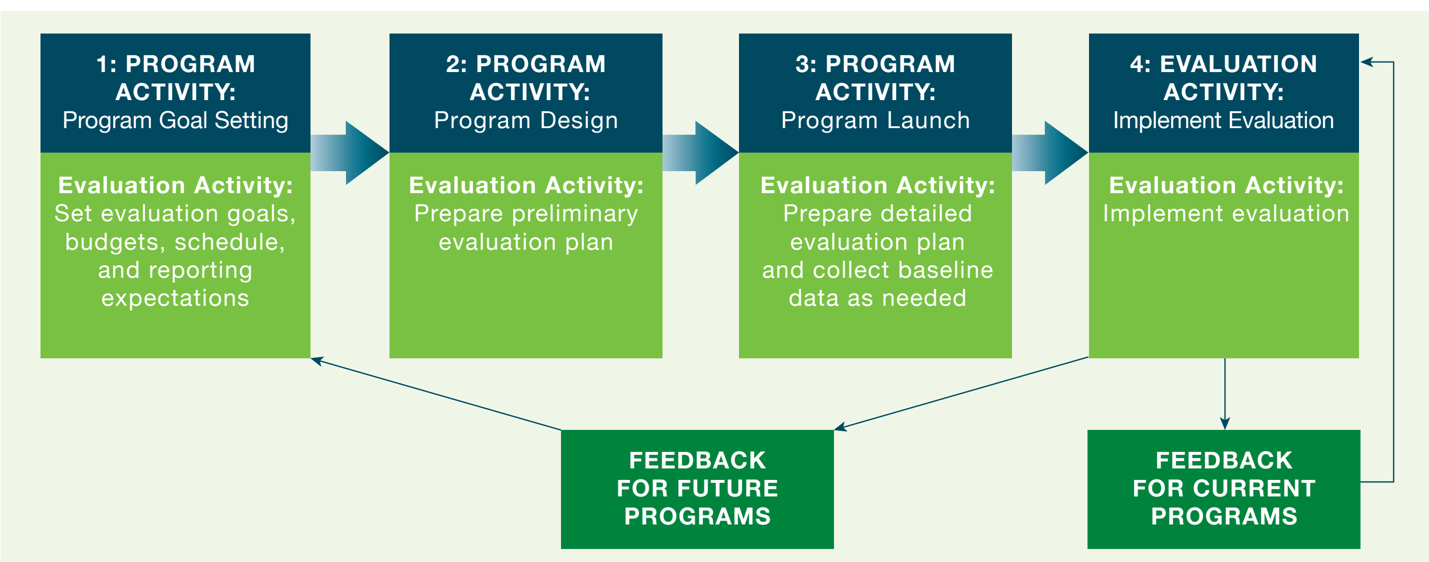
- **Program launch.** Program launch is when program materials and timing strategies are finalized and made ready, implementation contracts (if needed) are negotiated, trade allies and key stakeholders are notified, and materials and internal processes are developed to prepare for program introduction and implementation. If the detailed program evaluation plan has not been prepared, it 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 8.4.
- It is in the evaluation plan that the program-specific evaluation issues are fully addressed and resolved, including specifying the data needed to perform the evaluation as well as which data should be collected during program implementation. This is also the time when some baseline data collection can take place, but most likely only after the program's projects and participants are identified. In situations where project site baseline data collection is required as part of the evaluation, the evaluation planning process should be started well before a program is launched. Trying to collect baseline data on equipment that has been replaced (and usually thrown away) after a project is completed can usually be ranked between difficult and impossible.

The overall evaluation plan should be reviewed with program implementers and, equally important, with the appropriate oversight body(ies), to ensure that it meets the requirements of policymakers, portfolio managers, and/or regulators. These requirements are hopefully set in the evaluation framework, which is a high-level document prepared outside of the portfolio cycle as a guiding document for the relevant jurisdictions. This framework document concept is discussed in Section 8.2.

Although it often is preferable to start an evaluation prior to the program launch in order to collect baseline information, in most cases, either (1) the evaluation and program start simultaneously due to the typical interest in initiating a program as soon as possible, or (2) the evaluation starts well after the start of the program due to serial planning of programs first and then evaluation in the portfolio implementation cycle. In general, the sooner the evaluation process can begin, the higher the quality of the evaluation and the sooner information—such as indications of where programs are meeting, or not meeting, their goals—can be provided.

In terms of reporting, evaluation information can be summarized and provided on any time cycle, but most cycles are driven by (1) the time required to complete the evaluation work and prepare

FIGURE 8.1: Program implementation cycle with high-level evaluation activities



reports after the last projects in a program cycle are completed, (2) regulatory needs, and (3) a desire to get the information needed to implementers so they can adjust existing programs and design new ones (that may start right after the completion of the current program cycle) using current and relevant evaluation, measurement, and verification (EM&V) information. For future program designs, projected savings estimates can be adjusted based on program evaluation results. Assumptions underlying the energy efficiency potential analysis used for planning at the beginning of the program cycle can then be updated based on the full net impact analysis. These data then feed back into the goal setting and savings potential analysis activities, and the cycle repeats to allow for an integrated planning process for future programs.

8.2 HIERARCHY OF EVALUATION PLANNING AND REPORTING DOCUMENTS

There are several evaluation planning and reporting documents that are typically prepared as part of the evaluation effort. The following is an overview of these documents in the context of a portfolio of programs that uses public or utility customer funds and has a program administrator with some government (regulatory) oversight. While privately funded portfolios can use many of the concepts presented here, the formality and requirements for independent evaluation may be different. Publicly funded portfolios usually have more stringent reporting requirements than privately funded ones.

8.2.1 Evaluation Planning Documents

Entities such as states, regulatory commissions, utilities, and others can establish and document their evaluation requirements in a hierarchy of documents. Figure 8.2 outlines the hierarchy of these documents and indicates their typical time frame (or time horizon) and applicability level (e.g., state or utility program administrator, program or portfolio of programs, or individual projects). The increasing level of detail in this hierarchy of documents provides indications of the appropriate place for each stakeholder in the overall energy efficiency effort to provide input. For example, all stakeholders may be interested in an EM&V framework but—on the other end of the spectrum—only program implementers and evaluators will generally be concerned with the details of a program research or M&V plan.

Descriptions of the documents are as follows:

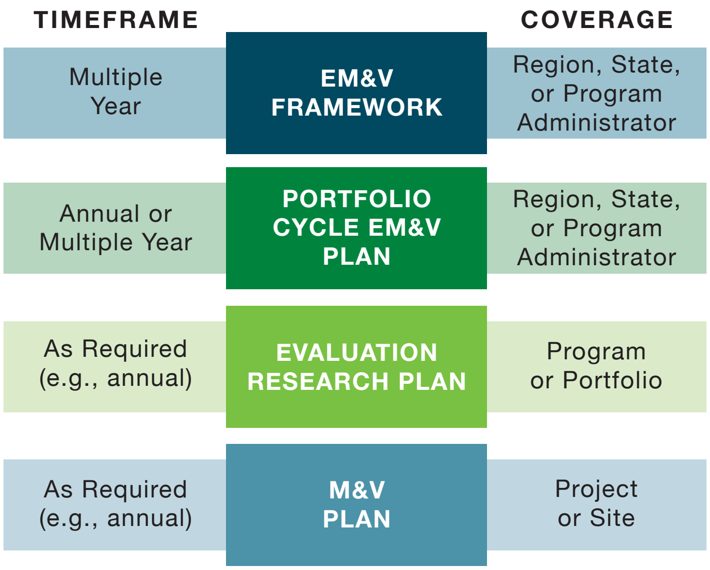
- **EM&V Framework.** A framework is a primary document that lays out EM&V principles, metrics, allowable approaches, net versus gross savings issues, reporting requirements and schedules, and the roles and responsibilities of various entities. An EM&V framework document tends to be "fixed" but can be updated periodically. It often sets the expectations for the content and scope of other EM&V documents (e.g., annual portfolio and statewide evaluation reports). The issues that would be considered when preparing such a framework are described in Section 8.3. This is perhaps the principle document that all stakeholders can focus on and use to provide high-level input—the "forest versus the trees" of an EM&V infrastructure.
- **Portfolio Cycle EM&V Plan.** This is the planning document that indicates which major evaluation activities will be conducted during the evaluation cycle (typically one, two, or three years) and describes them at a high level, including budget and allocation of resources and effort between programs, measures, market sectors, etc. It lists all the planned, major evaluation

EVALUABILITY

"Evaluability" is an assessment of the probability that evaluation information will be available when evaluations are actually undertaken. Some data (e.g., the age of a building) can be gathered at any time; some data (e.g. participant spillover, current hours of operation) are best gathered at the time of evaluation; 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 (these include 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 (projected) 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?

FIGURE 8.2: Hierarchy of EM&V planning documents



Source: Schiller, S. R.; Goldman, C.A. (August 2011). "Developing State and National Evaluation Infrastructures—Guidance for the Challenges and Opportunities of EM&V." 2011 International Energy Program Evaluation Conference Proceedings, August 16-18, 2011, Boston, Massachusetts. Madison, Wisconsin: International Energy Program Evaluation Conference (IEPEC). www.iepec.org/2011PapersTOC/papers/073.pdf#page=1.

activities with a schedule for when they are to be conducted. Examples of such major evaluation activities are the impact, process, and market evaluations, as well as updates to any technical reference manuals (TRMs). This EM&V plan might indicate, for example, (1) which programs would have rigorous evaluations in each year and which would only have verification reviews, and (2) which programs will undergo process evaluations and which will not.

- **Evaluation Activity-Specific Detailed Research Plans.** Research plans are created for the major EM&V activities or studies planned in a given cycle prior to the time each effort is launched. Examples of these plans are (1) program-specific impact evaluation plans that go into substantial detail on what evaluation approaches will be used, schedule, budgets, the data that will be collected, and results that will be reported; and (2) process, market effects, and market baseline study plans that would similarly provide sufficient detail to guide the actual implementation of the evaluation activity.

- **Project-Specific M&V Plans.** Project-specific plans may be required for projects that are part of a custom program and that are selected for analysis and inspection. These are the plans that describe the specific activities that would be conducted at a single site when one or more of the four International Performance Measurement and Verification Protocol (IPMVP) measurement and verification options are applied, or just for inspections if deemed savings values are to be used.

Jurisdictions, such as state or regional organizations overseeing energy efficiency portfolios, should also seriously consider developing TRMs. The contents of these TRMs vary, but they are typically a database of standardized, state- or region-specific algorithms (deemed calculations) and associated savings estimates (deemed savings values) for conventional electric and natural gas energy efficiency measures. The TRMs are used by energy efficiency program administrators and implementers to reduce EM&V costs and uncertainty in how measured savings will be credited with value as they prepare their projected and claimed savings values. Similarly, evaluators will also use the information in TRMs to prepare their evaluated savings values. TRM values for individual measures are not always formally vetted in a regulatory process, although this is a good practice. See Section 4.2 for more information on TRMs.

In some cases, different program administrators (e.g., utilities) in the same state or region will use different savings values for the same measures. Although different values may be perfectly appropriate for the same measures because of different baselines, delivery mechanisms, weather, or other factors, it is important for each jurisdiction to at least have established and consistent procedures for building and maintaining TRMs, with deemed calculations and savings values that can be accepted for projected, claimed, and evaluated savings estimates. State regulators and program administrators may also want to consider pooling their resources to support development of regional TRMs as well as statewide or regionally coordinated TRM updating processes and schedules. Examples of this are the Regional Technical Forum in the Northwest (www.nwcouncil.org/energy/rtf) and the Regional EM&V Forum in the northeast and mid-Atlantic regions (http://neep.org/emv-forum). Other TRMs available online are listed in Appendix C.

8.2.2 Evaluation Reports

As described in Chapter 3, there are impacts that are reported by the program administrators or implementers, and there are impacts reported by the evaluator. Savings estimates for programs tend to be reported in one or more of the following classifications:

- **Projected savings:** values reported by a program implementer or administrator prior to the time the energy efficiency activities are completed

- **Claimed savings:** values reported by a program implementer or administrator after the energy efficiency activities have been completed

- **Evaluated saving:** values reported by an independent third-party evaluator after the energy efficiency activities and impact evaluation have been completed.

Evaluators report the evaluation results and, as appropriate, work with (1) regulators to assess whether goals have been met, (2) program administrators to implement recommendations for current or future program improvements, and/or (3) resource planners to understand the historical role and future role of energy efficiency as an energy resource. Reporting also provides information to the energy consumers and general public, who are often funding these efforts, on what has been achieved with their investments.

Correlated with the evaluation planning documents described above, the following are the typical types of impact evaluation reports that are prepared.

- **Project-Specific M&V Reports.** These reports document the impacts determined for a specific project, site, or measure and the methods used to determine the impacts. They tend be reviewed with the project administrators and implementers before they are finalized, and their results are made available on a limited basis to protect the confidentiality of the consumer information that is usually included.
- **Impact Evaluation Reports.** The results of conducting the evaluation activities described in each impact evaluation plan are documented in an impact evaluation report. The report documents the impacts, and perhaps cost-effectiveness, of a program, as well as the methods used to determine the impacts. Program administrators and implementers usually have opportunities to provide input on these reports. The final reports are also usually publicly available.
- **Portfolio Cycle Evaluation Reports.** The results of carrying out the evaluation activities described in an EM&V portfolio plan are documented in a portfolio cycle (e.g., annual or biennial) evaluation report. It documents the impact metrics (e.g., gross and net energy and demand savings, first year, and lifetime) and cost effectiveness associated with the portfolio of programs as well as the methods used to determine the impacts. Program administrators and implementers also usually have opportunities to provide input on these reports. The final reports are made publicly available with summaries or synthesis reports provided in a manner that is accessible to laypersons and with guidance on context and interpretation of the evaluation findings. Interim reports are also suggested so that progress indicators can be provided and problems, if they exist, are identified before the end of the program cycle. Interim results can also support planning efforts for the next portfolio cycle, as it is common for the portfolio cycle evaluation reports to be completed well after the end of the completion of the evaluated cycle and into the time of the next cycle.

The above list indicates the typical reports associated with impact evaluations. However, other reports often prepared by evaluators include process evaluations and market assessments. These are prepared as required to support the implementation of effective energy efficiency programs. These reports are also made publicly available, preferably with summaries also provided in a manner that is accessible to laypersons.

8.3 PRINCIPLES AND ISSUES THAT DETERMINE THE SCOPE OF AN IMPACT EVALUATION

This section presents and discusses issues and principles that should be defined by each jurisdiction in order to develop a framework document for their evaluation activities. As with the last section, it provides an overview of these issues and principles in the context of a portfolio of programs that uses public or utility customer funds and has a program administrator with some government (regulatory) oversight. And, again, while privately funded portfolios can use many of the concepts presented here, the formality and typical requirements for independent evaluation may not be relevant.

Because of differences in approaches to EM&V and no "cross-border trading" of energy efficiency savings, it has not been necessary for regulatory commissions and administrators of ratepayer-funded energy efficiency programs in states to define a saved unit of energy in exactly the same way.[97] Thus, each jurisdiction (e.g., state) can, and many do, develop EM&V requirements (and TRMs) that are appropriate for their own situations, with the EM&V requirements for each jurisdiction linked to the needs of that jurisdiction.

How the evaluation issues and principles are defined for each jurisdiction depends on the specific programmatic and regulatory context (including any mandates) found within each jurisdiction, the objectives and scale of the energy efficiency activities being evaluated, and how EM&V results will be used. For example, one state may have very limited goals for energy efficiency and may not have performance incentives for their energy efficiency portfolio administrator. It also may have a limited level of naturally occurring or mandated (via codes and standards) energy efficiency activity. Another state may have established aggressive, long-term energy-savings targets in legislation, developed a performance-based incentives scheme for program administrators, and have high energy costs as well as a need for very solid savings data for resource planning purposes. The high energy cost scenario may also have resulted in a high level of natural and mandated energy efficiency activity. Given these differences, the first state may only need a limited level of EM&V and use fairly permissive baselines, while the second state might require very rigorous (and expensive) EM&V with well-defined baselines. Although in both scenarios EM&V can be valuable to regulators and program administrators and implementers, these kinds of considerations drive the type, style, budget, and timing of EM&V requirements.

These EM&V requirements can be developed by program administrators, agencies with responsibility for overseeing EM&V activities (e.g., state utility commissions, energy offices), and other stakeholders, and then documented in the framework document described in Section 8.2. Preferably, they all draw from established EM&V principles, such as those indicated in Section 7.6. To this end, stakeholders in each jurisdiction should work together to address the following 14 issues regarding EM&V. These issues below have been vetted in a number of jurisdictions and form a good basis for defining the needs of regulators, administrators, implementers, and resource planners. However, the list should be reviewed and customized for each jurisdiction, and to meet exigent regulatory needs.

1. **What are the evaluation objectives, metrics, and research issues that support program policies and/or regulations?** Driving the response to this question is the need to know the policy and/or regulatory goals that are the basis for the energy efficiency programs and thus the evaluation of the programs. See Section 8.3.1. and Section 2.1.1 for discussions of energy efficiency program policy contexts.
2. **What are the evaluation principles that drive the effort?** See Section 7.6 for a discussion of evaluation characteristics and principles.
3. **What is the scale of the evaluation effort?** How much time, effort, and money will be spent on evaluation? What is the relative allocation of resources among impact, market, and process evaluations? Section 7.5 includes a discussion on defining the scale of the effort and establishing related budgets.
4. **Who will conduct the evaluations, how is an independent evaluation defined, and what are the relative EM&V roles between implementers, evaluators, regulators, stakeholders, and others?** See the discussion below in Section 8.3.2.
5. **Is performance determined on the basis of net or gross savings?** What factors are included in defining net savings? Chapter 5 includes a discussion of net savings definitions, uses of net versus gross savings, and issues associated with determining and using net savings.
6. **What are the baselines against which savings are determined?** Section 7.1 includes a discussion of baselines and their influence on impact evaluations.
7. **What is the reporting "boundary"?** Are transmission and distribution (T&D) losses included, and how "granular" will the results be? See the discussion below in Section 8.3.3.
8. **What are the schedules for implementing the evaluation and reporting?** See the discussion below in Section 8.3.4.
9. **What impact evaluation approaches will be used?** This simply relates to selecting one or more of the impact evaluation approaches defined in Chapters 4, 5, and 6 and Appendix B (top-down evaluation) that will be allowed, or preferred, for the evaluation of different programs or the portfolio as a whole.
10. **What are expectations for savings determination certainty (confidence and precision)?** Section 7.4 includes a discussion of uncertainty and managing and defining the risk of savings uncertainty.
11. **Which cost-effectiveness tests will be used?** Cost-effectiveness test selection determines what data collection will be required as part of the impact evaluations. Cost-effectiveness tests are briefly described in Appendix B.
12. **How are evaluated savings estimates applied—looking back/going forward?** See the discussion below in Section 8.3.5.
13. **What are the data management strategies?** See the discussion below in Section 8.3.6.
14. **How are disputes addressed?** See the discussion below in Section 8.3.7.

These issues are presented in what can be considered a linear sequence, but many are interrelated and the overall planning process is certainly iterative. These issues can be addressed through a variety of mechanisms, such as collaborative efforts (as is the case in Washington), with advisory groups (as is the case in New Mexico), or with regulatory proceedings (as is the case in many states). The end result of addressing the above 14 issues is the evaluation framework that documents the evaluation expectations for all stakeholders. Experience has indicated that, if the funding and time requirements for reliable evaluations are understood and balanced with information needs and accuracy expectations and approved by the primary stakeholders, evaluation efforts can be well-supported and succeed in providing the results desired.

8.3.1 Defining Evaluation Objectives and Metrics

Impact evaluations should focus on a program's performance at meeting its key implementation goals and, if desired, provide information for future program and resource 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. This requires consideration of the specific program benefits (metrics) to be evaluated and reported, the concepts behind the results the program is expecting to achieve (usually referred to as the program theory), and whether any other related evaluations will be concurrently conducted and coordinated.

Under-designed evaluations can waste valuable resources by not reliably providing the information needed—or worse, providing incorrect information. Related to under-designed evaluations are delays that 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. Evaluation activities should be prioritized so that evaluation resources—typically limited—can be focused on the issues of importance. Like most activities, an evaluation that is well defined in terms of objectives and scale, and also affordable, is more likely to be completed successfully than one with undefined or unrealistic objectives and budget requirements.

Setting impact evaluation objectives involves defining the specific information (metrics) that will be reported from the evaluation. The scale of the evaluation is a more general concept, indicating how much effort (e.g., time, funding, human resources) will be expended on the evaluation. It involves balancing the various needs for information, accuracy of the information, and timing of the information against the available resources, including funding.

8.3.1.1 Evaluation Objectives

As discussed in the beginning of this guide, evaluations can have three primary objectives:

1. To document the benefits/impacts of a program and determine whether the program (or portfolio of programs) met its goals
2. To help understand why program-induced effects occurred and identify ways to improve current and future programs
3. To support energy demand forecasting and resource planning by understanding the historical and future resource contributions of energy efficiency compared to other energy resources.

Therefore, the first step in planning an evaluation is simply picking which of these objectives (if not all) are applicable, prioritizing them, and making them more specific to the evaluated program. The following are some typical and more specific impact evaluation objectives:

- Measure and document energy and peak savings
- Document program milestones, such as homes weatherized or people trained
- 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)
- Assess if there is a continuing need for the program
- Provide specific data for demand forecasts and resource planning in an integrated resource planning (IRP) effort (in the format and with baselines and metric definitions consistent with the resource planners' requirements).

In practice, the selection of evaluation objectives will be shaped by many situational factors. Probably the most important factors are the program goals—hopefully goals that are well defined and quantifiable into evaluation metrics. Some program goals, beyond numerical metrics of energy and demand savings, cost effectiveness, and maximizing energy or peak savings within portfolio budgets, are as follows:

- Maximize leverage of portfolio dollars in creating private investment in energy-efficient products to achieve savings
- Defer specific resource planning needs (e.g., peaking or baseload power plants, transmission or distribution investments)
- Reduce greenhouse gas (GHG) emissions
- Maximize the fraction of the population participating in portfolios by sector
- Maximize consumer satisfaction.

In Section 2.1.1, three common examples of program goals are briefly described. For those three examples, the following are some comments on evaluation objectives:

- **Cost effectiveness.** All cost-effective energy efficiency evaluations for programs with this policy context have to focus resources on determining the long-term cost effectiveness of the programs. Evaluation methodologies can be complex, as assumptions about the baselines, net-to-gross ratios, and non-energy benefit factors used in program evaluations can be critical to the calculations selected for determining cost-effectiveness.
- **Energy efficiency resource standard (EERS).** Evaluations for programs with an EERS policy objective may be able to focus relatively less effort on long-term cost effectiveness, because the EERS mandates are often promulgated on the assumption (which can only be changed by future legislation) that energy efficiency will always be less expensive than other resources up to the level of the EERS target percentage. Evaluations of EERS programs thus focus on the year-by-year tally of progress toward the EERS targets.
- **Target spending budget.** Evaluations for programs with target budgets have to focus on the most cost-effective use of the

budgets, so that they can demonstrate that the administrators are getting the maximum benefit from each program dollar.

Another policy issue that can affect program evaluation budgets is whether the utility or other program administrators are eligible to receive a financial incentive (bonus) for managing a successful program portfolio—or suffer a penalty for not meeting the established goals. The scale of the penalty or bonus can determine the level of rigor required in the evaluation as well as the level of attention paid to the evaluation activities. Evaluators should be cognizant of the level of intensity that some stakeholders apply to seemingly obscure issues of evaluation methodology when they can affect program goals and consequences, such as administrator incentives.

However, beyond just what the program goals are and uses of the evaluation results, other factors also affect the selection of evaluation objectives:

- Whether the program is new, expanding, or contracting
- 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).

One approach to address these considerations is to use theory-based evaluation (TBE). In short, TBE involves defining the events a program is intended to cause, along with the precise causal mechanisms leading to these events, with the evaluation, then focusing on testing the consistency of observed events with the overall program theory (and thus the goal). Theory-based evaluation is briefly discussed in Section 7.10.3, in the market transformation program section.

8.3.1.2 Evaluation Metrics

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

- **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 the measure's lifetime, in years
- **Monetary savings:** dollars saved in electricity or natural gas costs and maintenance per year.

The metrics listed above are related to energy savings and avoided emissions. However, other non-energy metrics may also be a subject of the impact evaluation; examples include direct or indirect job gains created over the program implementation period for a training program or annual water savings associated with a low-flow faucet program that is focused on water heating savings. In addition, a focus of many programs is transforming markets (see Section 7.10.3 and Appendix B), and thus other metrics can include market penetration of energy efficiency products or pricing of such products.

As a reminder, as discussed in Section 7.4, evaluation results, like any estimate, should be reported as expected values with an associated level of variability.

8.3.1.3 Other Evaluation Efforts and Other Programs

While this guide is focused on impact evaluations, there are other types of evaluations (as described in 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 objectives and 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 the programs, their evaluations, and the assignment of benefits to one program versus another need to be coordinated to avoid—or at least minimize—double-counting of savings.

8.3.2 Evaluator Roles and Selection

While evaluation is relevant to many kinds of business and government efficiency portfolios, the importance of defining the role and independence of the evaluator is primarily an issue for programs that are funded with public or energy consumer funds. In such programs, a government regulatory entity, typically a utilities commission, oversees the actions of the program administrator. This subsection addresses the situation in which a government regulator oversees a program administrator's conduct of efficiency activities and wants some level of independent assessment of program/portfolio impacts. It first describes the possible roles of evaluators in program implementation and evaluation, and then presents criteria for defining independent third-party evaluators and criteria for their selection.

Starting with impact evaluation—particularly determining energy and demand savings—there are two possible sets of results reported after a project or program is implemented: claimed savings and evaluated savings. Claimed savings, those reported by the program administrator or implementer, are needed not only for possible reporting to a government/regulatory agency that oversees efficiency programs (such as utility program administrator reporting to a utilities commission) but also for the basic implementation of the programs. For example, on a project-by-project basis, claimed savings need to be determined in order to pay contractors or consumers whose payments depend on actual implementation and/or achieved savings. Claimed savings also need to be reported internally to justify expenditures, as would be the case with any organization. Staff or consultant evaluators (e.g., engineers, analysts, econometricians), working directly for the program implementers and/or administrators, almost always prepare these claimed savings. Their role is to directly support the program implementation and prepare required internal and external (e.g., regulatory) reporting.

For the above-listed administrator/implementer functions, those doing the work may or may not have the formal title of evaluators. Evaluator is a very broad term that can describe people conducting evaluation activities who are part of a consulting, program administration, implementation, government, utility, or other type of organization. If not called evaluators, they may be known as analysts, engineers, or M&V engineers, or may simply be part of the administrator's or implementer's program implementation team.

Evaluated savings, on the other hand, are only required if some entity, such as a government regulatory agency, wants an independent third party to either determine themselves or double-check the claimed savings that are determined by the implementer/administrator. This leads to several resulting questions:

- Who is responsible for which evaluation activities?
- Will there be overlap such that an independent third-party evaluator, in preparing evaluated savings, only simply confirms the work done by the implementer/administrator in preparing the claimed savings (verification)? Or, will the independent third-party evaluator conduct some of its own data collection and impact evaluation analyses?
- What is meant by an independent third-party evaluator, and who retains them to prepare the evaluated savings reports?

Starting with the last question first, there is no formal definition of independent or third-party evaluator, as well as there are no well well-established precedents as to who hires the entity(ies) that provides the evaluated savings reports. The hiring entity could be the regulator or the administrator, or perhaps some other entity. However, in general practice, "independent third party" is thought to mean that the evaluator has no financial stake in the evaluation results (e.g., magnitude of savings) and that its organization, its contracts, and its business relationships do not create bias in favor of or opposed to the interests of the administrator, implementers, program participants, or other stakeholders such as utility customers (consumers). However, different states' regulatory bodies have taken different approaches to (1) defining the requirements for evaluators who are asked to review the claimed savings and prepare evaluated savings reports, and (2) who hires that evaluator.

A 2012 American Council for an Energy-Efficient Economy (ACEEE) study indicated the following:[98]

> ….. there is a great diversity among the states in how they handle the evaluation of ratepayer-funded energy efficiency programs. This begins with the administration of the evaluation function itself, where just over a third of states (37%) feature utility [evaluation] administration, 36% feature [evaluation] administration by the utility regulatory commission or a combination of the commission and utilities, and over a quarter (27%) feature [evaluation] administration by some other government agency or third-party entity. Most states (79%) rely on independent consultants/contractors to conduct the actual evaluations, although a substantial minority (21%) use utility and/or government agency staff.

Irrespective of how the relationships are determined or who hires whom, the objective is for all parties to the evaluation to believe that the reported results are based on valid, unbiased information that is sufficiently reliable to serve as the basis for informed decisions.

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 at the tail end of the project to assess delivery, cost-effectiveness, and achievement of stated goals. 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 timely assessment and recommendations, if needed, for corrective actions. 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."

In terms of the first and second questions listed above—who does what and whether there are overlaps—there are many options, as alluded to in the ACEEE study referenced above. Having complete analyses conducted independently by a third party provides perhaps the greatest level of due diligence and integrity for evaluated savings values. However, such analyses do add costs, and it is not uncommon for there to be overlaps in the determination of savings. Also, because a common objective of evaluation is to improve the performance of a program and help with program improvement, a totally independent approach does not directly favor a tight working relationship between the evaluator and the implementer/administrator. Thus, the selection of an evaluator can require balancing evaluation independence (so that the evaluation is considered objective) with the desire to have the evaluator close enough to the process that the evaluation provides ongoing and early feedback without the implementer feeling "defensive."

One way to look at the relative roles of the different entities involved in preparing claimed and evaluated savings (as well as project savings) is to consider the roles as associated with oversight or administrator activities. These can be generally defined as follows:

- **Oversight activities:** activities that are under the purview of the entity responsible for all energy efficiency programs and the associated EM&V implemented by program administrators in the subject jurisdiction (e.g., state or utility service territory). Oversight activities will usually include coordination with the government/regulatory authority. They may also include feedback or guidance to and from stakeholders (including administrators and implementers) about the evaluation plans and implementation as well as the process of approving reported results.
- **Administrator activities:** activities undertaken by the program administrators during the process of developing, implementing, and conducting M&V activities pertinent to their implementation of energy efficiency programs. These M&V activities may also be known as the *primary evaluation activities*.

As suggested in Section 8.2, it is in the evaluation framework that these roles and responsibilities are sorted out. One example is how Maryland defined these roles in a simple matrix included in Table 8.1 at the end of this chapter. This is only an example, however, not necessarily a recommended allocation of duties.

In terms of selecting third-party independent evaluators, the usual approach is via a request for proposal (RFP) or request for qualifications (RFQ) process. Typical selection criteria include qualifications and experience of particular staff assigned to the proposed work, resources that the bidding organization (or often, bidding team) bring to the project, availability of the key staff, labor rates and overall budget expectation, and approach to the impact evaluation work, including schedule and deliverables.

Budget expectation and approach to the work (instead of a set work scope, deliverables, and budget) are listed in these criteria because it is usually not practical to ask bidders to prepare a complete scope/budget during the proposal process. This is because either the programs to be evaluated are not fully defined and/or the typical first task is to prepare the portfolio cycle (e.g., annual or biennial) EM&V plan and the evaluation activity-specific detailed research plans, if not the evaluation framework itself.

Traditionally, evaluation consulting firms tend to use econometricians (professionals who apply statistical and mathematical techniques to problem solving) and engineers. In the last 10 years or so, there has been acknowledgement of the need for individuals trained in fields other than just efficiency and/or with other skills, such as market analysis and public policy. People with skills such as interviewing and survey data analysis are also needed.

Many evaluators are members of industry professional organizations or are Certified Measurement and Verification Professionals (CMVPs).[99] The following are two of the professional organizations that energy evaluators participate in (and which will post evaluation RFPs/RFQs):

- Association of Energy Service Professionals, www.aesp.org
- International Energy Program Evaluation Conference, www.iepec.org.

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

8.3.3 Setting the Boundary and Granularity of Reported Results

When evaluating energy, demand, and emissions savings, it is important to properly define the project boundaries (i.e., the equipment, systems, or facilities that will be included in the analyses). Ideally, all primary effects (the intended savings) and secondary effects (unintended positive or negative effects—sometime called *interactive factors*),[100] and all direct (at the project site) and indirect (at other sites) effects will be taken into account. From a practical point of view, with respect to energy and demand savings, this translates into deciding whether savings will be evaluated for specific pieces of equipment (where the "boundary" may include, for example, just motor savings or light bulb savings), the end-use system (such as the HVAC or the lighting system), whole facilities, or an entire energy supply and distribution "system."

From an electricity system point of view, the energy and demand savings at the power plant producing the electricity will be greater than the savings at the end-use (in the facility) due to transmission and distribution (T&D) losses. These average or marginal (which can be much higher than the average) losses are on the order of 5% to 20%.[101] Thus, the savings at the power plant busbar can be 120% of the end-use savings. Whether T&D losses will be included in the boundary and reported as part of program impacts is something that can be defined in the evaluation framework.

For avoided emissions calculations, the boundary assessment issues are discussed in Chapter 6. Boundaries are also important for defining other non-energy benefits (NEBs) as well. For example, if there is job impact to be reviewed, it needs to be decided whether the jobs are local, statewide, national, or international.

The time granularity of evaluation analyses relates to whether 15-minute, hourly, monthly, seasonal, annual, and/or lifetime data collection and savings reporting are required. The "granularity decision" is based on how the information from the evaluation is to be used. Annual savings data are generally only useful for an overview of the program benefits. More detailed data are usually required for cost-effectiveness analyses, demand forecasting, and resource planning. 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) is related to time granularity of the evaluation results. Section 7.2 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.

8.3.4 Schedule for Evaluation and Reporting

The evaluation time frame has two components: the time period over which the evaluation activities will take place and the reporting schedule. An ideal evaluation schedule begins before the start of the program implementation (to collect any baseline data and set up the overall evaluation infrastructure) and continues 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.

The first consideration is when to start the evaluation efforts. Programs and portfolios tend to get into a regular cycle; for example, the start of each program year may be on January 1, but the evaluation process may not get started until the spring. One approach in this situation is to accept the late start of an evaluation for the current year and conduct a more limited effort than desired, but then move the evaluation cycle and the program/portfolio cycle into better alignment within one or two years. Other evaluation activity schedule considerations are the length of the portfolio cycle (one, two, or three years), whether to estimate persistence of savings (see Section 7.3), whether information is needed for a next portfolio cycle, and what information will be included in the evaluation reports.

In terms of reporting requirements, certainly there will be a need for final program and portfolio reports for each program year and/or cycle. However, there will likely also be needs for interim reports to track the performance of the programs being evaluated and the evaluation itself.

There are several considerations for setting up a reporting schedule:

- The timing for policy decisions and evaluation planning
- The desire to have early feedback for program implementers
- Program lifecycle stage (evaluating a first-time program or a long-established program)
- Evaluation data collection time lags
- Regulatory and/or management oversight requirements
- Contract requirements for reporting savings for "pay for performance" programs
- Timing requirements to use the evaluation results to update energy and demand savings as well as measure life estimates for specific energy efficiency measures in a TRM.

A standard reporting plan would have final reports for each of the project-specific M&V reports, impact evaluation reports, and portfolio cycle evaluation reports, with interim reporting on the impact and portfolio cycle reports. Quarterly or biannual evaluation status reports can also be helpful for keeping all stakeholders informed. In addition, providing online summary data to the public can be a valuable tool for keeping publicly funded efforts informed and perhaps for maintaining public support.

8.3.5 How Are Evaluated Savings Estimates Applied—Looking Back Or Going Forward?

Estimates of costs and savings from energy efficiency measures are typically made both prior to program implementation (i.e., projected savings) and post-program implementation (i.e., evaluated and/or claimed savings). And, as one would expect, evaluated estimates of savings are considered a more accurate representation of actual savings than projected savings. An issue arises when deemed savings values are used to project and claim energy savings for an energy efficiency measure in a given program year (e.g., based on per-unit savings values in a TRM approved for that program year), but an evaluation during the program year indicates the TRM per-unit savings values are too high or too low for the subject measure. The question thus becomes "Should the deemed savings value be adjusted retroactively to the current program year or only applied on a going-forward basis?"

Consider the following example:

- TRM per-unit savings values developed in 2011 indicate that the savings from measures verified to have been installed in a program in 2012 are 10,000 MWh.
- However, an ex post (after the fact) evaluation indicates that the per-unit savings values in the TRM were overly optimistic, and the actual program savings are 9,500 MWh based on the same number of units verified to have been installed.
- Assuming the verification is correct and all the measures were installed, does the oversight body (e.g., a regulatory commission) credit the program with 10,000 MWh of savings or only 9,500 MWh for the subject program year?

There are equity issues associated with the above options for how and when to apply updated TRM values. On one hand, the program administrator/implementer relied on an approved value for budgeting and savings estimates; however, for resource planners and other stakeholders, what matters most is the most accurate indication of what occurred. While no perfect solution exists, this is an issue that is best to decide upon for inclusion in an evaluation framework before it occurs during the portfolio cycle. One approach is to use this guidance:

- Per-unit cost and savings stipulated (deemed) values in a TRM should be based on the best available information at the time these estimates and/or calculations are made, and they should be determined in a rigorous and transparent manner. It is recognized that TRM values are not static and will be updated from time to time.
- If new information indicates that per-unit cost and savings stipulated (deemed) values in a TRM should be updated, these new values will be adopted for use in future program savings projections, claims, and evaluations.
- Savings claimed in the current program year (based on deemed savings) for measures or programs already implemented before the TRM is updated are not adjusted retroactively up or down for purposes of defining program administrator or implementer goal achievement.
- In terms of using the evaluated results for load forecasting and resource planning, estimates with the most up-to-date TRM values are used, although this could result in two sets of results being reported: one for forecasting/planning and one for administrator goal achievement. However, once TRMs become "stable" (i.e., fewer and fewer changes are made) the difference should be minor, if there is any at all.
- Savings from custom projects or programs (where savings are determined after project implementation using agreed-upon protocols) should use the evaluated values as the savings for all purposes.

8.3.6 Data Management Strategies

Evaluations are based on data—often in very large amounts. Decision makers often do not need or want to see the detailed calculations or raw data inputs that drive evaluation results. But all parties need to know that reported results are built on a foundation of sound data inputs. Proper data management strategies allow administrators, implementers, and evaluators—as well as oversight bodies—to delve into the underlying data, both to be able to review underlying assumptions and to combine the data in new ways as they see fit for current program reviews or future program developments.

Thus, while data management could easily be defined as "down in the weeds," it can be a major attribute of a portfolio's implementation and evaluation and a major cost. Thus, it deserves attention in the evaluation framework. For stakeholders, the questions addressed in the framework tend to be associated with: "Will the data resources be fully used for current and future project development, such as updating deemed savings values?" "Will consumer confidentiality be properly maintained?" and "What will be publicly accessible?"

The fundamentals of good data management are the same across industries. Within an energy efficiency portfolio, two areas of primary importance are the ability to compare results across time (longitudinal analysis) and ability to compare results by factors such as program type or delivery mechanism. Some items are typically specified as part of an evaluation plan:

- Data required
- Format of data to be provided by tracking systems; compatibility and standardization
- Access to data and summaries
- Data confidentiality protection protocols
- Data quality assurance and control.

In terms of ensuring consistency of data reporting across programs, the following data components are recommended for standardization:

- Measure naming convention (i.e., the same energy measures, end-uses, and applications have the same name from program to program)

- Measure categorization (i.e., place measures into logical measure categories, subcategories, and technologies: Category = Lighting, Subcategory = Indoor Lighting, Technology = CFL)
- Sector and building type classification across programs (e.g., Sector = Commercial, Building Type = Education—Primary School).
- Normalization units across programs (i.e., utilize normalization units as the basis of quantity or measure size, such Flow-GPM for water flow measures or Ctrl-kW for occupancy sensors).
- Program delivery methods across programs (e.g., program types are prescriptive rebates, direct install, and point of sale)
- Indication of project dates (e.g., date of installation).

While there is not currently a single reporting standard for energy efficiency programs, national, regional, and international reporting specifications are under development. In the meantime, three database projects that demonstrate good data management practices are as follows:

- **Standardized Project Tracking (SPT) database in California.** California has long maintained various databases related to energy use and energy efficiency. In recent years, the California Public Utilities Commission has worked closely with the program administrators to create standard claims-reporting templates. Recent versions of the SPT allow all stakeholders in California to review program savings and costs at the measure, program, and portfolio levels. See the websites at http://open-emv.com/ and www.cpuc.ca.gov/PUC/energy/Energy+Efficiency.
- **The Regional Energy Efficiency Database (REED).** Developed by the Regional EM&V Forum, REED has commitments from ten jurisdictions for a common reporting system of energy efficiency savings and associated impacts to support a range of energy and environmental policies. See the website at http://neep.org/emv-forum.
- **The Northwest Regional Technical Forum (RTF).** The RTF is an advisory committee established in 1999 to develop standards to verify and evaluate conservation savings. See the website at www.nwcouncil.org/energy/rtf/Default.htm.

8.3.7 Addressing Disputes

Disputes can arise from the process used to develop impact evaluation results and/or the results themselves. Disputes are best addressed before they arise through understood agreements on how the evaluations will be conducted (i.e., defined in an evaluation framework and through good communication). However, disputes can and do arise, and it is best to define how they will be addressed before they occur. Most jurisdictions will have their own approaches, with mechanisms for regular discussions, regulatory hearings, mediations, arbitration, or other solutions. Even a few lines in an evaluation framework document defining the steps for dispute resolution can eliminate a great deal of difficulty should a dispute arise.

8.4 PORTFOLIO CYCLE IMPACT EVALUATION AND M&V PLAN OUTLINES

The evaluation planning documents should clearly present the evaluation efforts and details of the actions to be undertaken during the evaluation activity, as well as consideration of regulatory (reporting) requirements. A 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, why, when, and how. Plans are also important historical documents since it is not unusual for programs with long lifecycles to undergo staff changes over the course of the program.

The following subsections outline the contents of a portfolio cycle impact evaluation plan and 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 is an evaluation planning checklist.

8.4.1 Portfolio Cycle Impact Evaluation Plan and Report Outlines

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

Part A. Program Background

1. Short description of the program(s) being evaluated (e.g., the market, program delivery approach, technologies, budget, objectives)
2. Presentation of how the program will achieve its objectives, the program theory
3. List of the technologies offered by the program
4. Program schedule
5. Numerical energy and non-energy savings projections.

Part B. Evaluation Overview

1. List of evaluation objectives and how they support program goals
2. List of which metrics will be reported (e.g., annual MWh, monthly peak kW, annual therms, annual CO_2)
3. Description of verification activities
4. Version of the TRM to be used and/or any TRM development/review activities

5. Gross and net impact evaluation approaches selected for determining energy (and demand) savings
6. Methodology for calculating non-energy benefits such as avoided emissions, as appropriate
7. List of primary factors that will be considered in analysis of gross and net savings (e.g., weather, occupancy, free riders, spillover) as well as list of major assumptions
8. Description of how program impact results will be combined to report portfolio impacts, addressing the need for adjustments such as accounting for program overlap or other factors
9. Expectations for overall certainty of savings estimates
10. Assumptions concerning availability of data and other information provided by the administrator/implementer; relative roles of evaluator and administrator/implementer
11. Budget and schedule summary
12. Listing of evaluators (if known) or evaluator selection method.

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

(This is the detailed presentation of evaluation activities to be undertaken, including the evaluation approach to be used.)

1. Gross impact savings analysis—a description of the data collection and 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 (as appropriate)
3. Data collection, handling, and sampling:
    a. Measurement collection techniques
    b. Sampling approach and sample selection methods for each evaluation activity that includes sampling efforts (as appropriate)
    c. How the comparison group or non-participant information will be used in the evaluation(s) and in the analysis (as appropriate)
    d. 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, including reporting activity and key milestones, including communications with administrator/implementer
6. Detailed budget and schedule
7. Evaluation team—information concerning the independence of the evaluator.

The final product or output of an evaluation is an evaluated savings report. The following is a sample report outline:

- 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
    - Report Structure
3. Study Methodology
    - Data Collection Approach(es)
    - Analysis Methods
    - Limitations, Caveats
4. Key Evaluation Results with metrics and realization rates (answers for all of the questions specified for the evaluation. This 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. Synthesis of analysis and findings as well as implications of findings
6. Recommendations (should include clear, actionable, and prioritized recommendations that are supported by the analysis)
7. Summary and Conclusions
8. Appendices (examples listed below):
    - Recommended improvements to the evaluation process, including any lessons learned for future evaluation studies
    - Detailed documentation of the research design and assumptions, data collection methods, evaluation analysis methodology, references, and results tables
    - 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 to results
- Possible sources of overestimation and underestimation
- Analysis of reliability of energy savings estimates, treatment of issues that threaten reliability of results (e.g., double counting, use of savings factors, and synergistic effects)
- How attribution was addressed (for net savings impact evaluation)
- Other assumptions and justifications.

8.4.2 Project-Specific M&V Plan and Report Outlines

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

M&V activities involve the following steps:

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, and 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 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, determining current year savings, identifying factors that may adversely affect savings in the future, and estimating savings for subsequent years.

A project-specific M&V plan describes in reasonable detail what will be done to document project savings. It can be a plan for each energy efficiency measure included in the project (e.g., when a retrofit isolation approach is used). Or, it can cover the entire project (e.g., when the whole-facility analysis approach is used). In either case, the M&V plan will consider the type of energy efficiency measures involved and the desired level of savings certainty.

The M&V plan should include an overall project description, facility equipment inventories, descriptions of the proposed measures, energy-savings estimates, a budget, and proposed project implementation and M&V schedules. A project-specific M&V plan should also 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. elected 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 in the analyses
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
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 tables at the end of this chapter summarize what could be contained in the M&V plans. Table 8.2 lists general requirements for an overall M&V plan. Table 8.3 lists requirements that could be addressed in the M&V plan 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 (www.evo-world.org).

8.4.3 Checklist of Planning Decisions for an Impact Evaluation

Table 8.4 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 8.1: Maryland Public Service Commission EM&V, Data Tracking, and Reporting Act Roles and Responsibilities for Utilities and Statewide Evaluation Consultants

TASK AND/OR DELIVERABLE	UTILITIES (OR THEIR CONSULTANT)	PSC STATEWIDE EVALUATION CONSULTANT(S)	MD PSC
STATEWIDE STUDIES			
Prepare statewide baseline study		X[102]	
Prepare additional statewide market assessments (e.g., market impact studies) and updates (bi- or tri-annual)		X	
Prepare statewide technical reference manual (TRM) (and annual or bi-annual updates)	X[103]		
Review TRM and updates of TRM		X	
Approve TRM (and review statewide market assessments)			X
PLANNING AND MANAGEMENT			
Prepare Master Evaluation Schedule for utility and statewide evaluation contractor activities; include overview of reporting schedule with annual and tri-annual portfolio reporting, as well as semi-annual interim reports and/or presentations	X	X	X
Develop utility (statewide and individual utility) impact and process evaluation plans; including gross and net energy and demand (including peak demand from distributed resources) savings, cost-effectiveness analyses, database and reporting protocols, survey templates, and schedules	X		
Review utility evaluation plans		X	
Develop plan for due diligence (quality assurance/quality control, QA/QC) of utility impact results (energy and demand savings, cost-effectiveness), including verification approach (with sampling plan) and schedules for review of utility submittals and reporting to PSC		X	

TASK AND/OR DELIVERABLE	UTILITIES (OR THEIR CONSULTANT)	PSC STATEWIDE EVALUATION CONSULTANT(S)	MD PSC
PLANNING AND MANAGEMENT (CONTINUED)			
Develop plan for determining and reporting additional outcomes (e.g., system reliability,; T&D and generation needs impacts; emissions avoidance; price mitigation; jobs impacts; effects on ratepayers, especially low-income; coordination with federal stimulus funding)		X	X
Review and approve utility and PSC evaluation consultant plans			X
Coordinate all utility evaluation efforts	X		
Coordinate statewide due-diligence (QA/QC) impact evaluation/ verification efforts		X	
PROCESS EVALUATION			
Prepare program process evaluations	X		
Conduct (independent) customer and trade-ally satisfaction surveys and reports	X		
Review customer and trade-ally satisfaction survey results		X	X
DATABASES			
Prepare a data reporting, interface, and database plan that includes coordination between the utilities' and PSC's statewide evaluation contractor database(s) and utility databases	X	X	
Design, implement, and maintain utility primary program management and tracking database(s) with project and program data (includes individual utility databases and a statewide database operated by utilities)	X		
Design, implement, and maintain statewide data management and quality control database of information "uploaded" from utility database(s) and used for (1) obtaining and managing data for due-diligence activities, and (2) establishing a public web-accessible database and reporting system with aggregated, higher-level information on program impacts (e.g., statewide energy and demand savings)		X	
Review and approve statewide database and reporting plan			X

TASK AND/OR DELIVERABLE	UTILITIES (OR THEIR CONSULTANT)	PSC STATEWIDE EVALUATION CONSULTANT(S)	MD PSC
PRIMARY DATA COLLECTION AND IMPACT ANALYSES: ENERGY EFFICIENCY (EE) AND DEMAND RESPONSE (DR) PROGRAMS			
Prepare ex ante (preliminary) savings estimates	X		
Conduct primary data collection and site baseline and ex post (after the fact) verifications for energy efficiency and DR projects	X		
Prepare persistence of savings analysis: Conduct primary data collection	X		
Prepare analyses and documentation of project, program, and portfolio gross and net energy and demand savings, cost effectiveness	X		
INDEPENDENT DATA COLLECTION AND IMPACT ANALYSES			
Conduct quality control and due diligence of utility analyses and documentation of project, program, and portfolio gross and net energy and demand savings and cost effectiveness; inspect sample of project sites and review primary data and analyses, prepare verified achieved versus claimed savings and cost-effectiveness report per reporting schedule		X	
OTHER OUTCOME ANALYSES			
Prepare additional efficiency and DR program/portfolio outcome reports: System reliability, T&D and generation needs impacts; emissions avoidance; price mitigation; jobs creation; impacts on low-income ratepayers; and leveraging of federal stimulus funding such as American Recovery and Reinvestment Act (ARRA) funds		X	
REPORTING			
Prepare utility interim semi-annual and final annual (and tri-annual) reports of energy efficiency program and portfolio net and gross impacts and cost-effectiveness evaluation results	X		
Prepare semi-annual (interim) and annual (final) reports of verified achieved utility program and portfolio results for energy and demand savings and cost effectiveness.		X	
Prepare annual report on additional Empower MD outcome results (e.g., avoided emissions, reliability and job impacts)		X	
Review utility and PSC consultant semi-annual reports; review and approve utility and PSC consultant annual (and tri-annual) reports			X

TASK AND/OR DELIVERABLE	UTILITIES (OR THEIR CONSULTANT)	PSC STATEWIDE EVALUATION CONSULTANT(S)	MD PSC
BEST PRACTICES			
Participate in quarterly (or semi-annual) impact evaluation process review and improvement meetings	X	X	X
Prepare best practices recommendations for improvements to evaluation processes	X	X	
Review and approve best practices recommendations for program modifications and improvements			X
OTHER			
Prepare materials and reports in support of PSC analysis of efficiency programs		X	X
Organize and conduct periodic statewide public workshops on evaluation results of energy efficiency programs		X	X

Source: Consensus Report on the EmPower Maryland EM&V Process. (June 24, 2009). .

TABLE 8.2: Energy Efficiency M&V Plan Contents: General Components

CATEGORY	M&V PLAN COMPONENTS
Project Description	Project goals and objectives
	Site characteristics and data available (e.g., whole building and end-use metering)
	Measure descriptions that include how savings will be achieved
Project Savings and Costs	Estimated savings by measure
	Estimated M&V cost by measure
Schedules	Project installation schedule
	M&V activities
Reporting	Raw and compiled data formats
	M&V report contents
	Reporting interval
M&V Approach	Confidence and precision of results
	M&V option(s) used
	Person(s) responsible for M&V activities

TABLE 8.3: Energy Efficiency Project-Specific and Measure-Specific M&V Plan Contents

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 calculated equations	kWh $savings_{(year)}$ = [(kW/$Fixture_{(baseline)}$ x $Quantity_{(baseline)}$) – (kW/$Fixture_{(post)}$ x $Quantity_{(post)}$] x Operating Hours per $year_{(post)}$
	Regression expressions	Three parameter change-point cooling model
	Computer simulation models	DOE-2 simulation model
Metering and Monitoring	Metering protocols	ASHRAE Guideline 14-2002 pump multiple point test throughout short-term monitoring
	Equipment	ABC Watt Hour Meter
	Equipment calibration protocols	National Institute of Science and Technology protocols
	Metered data	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

TABLE 8.4: Evaluation Planning Checklist

CHECKLIST FOR GROSS SAVINGS DETERMINATION	
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	
Randomized controlled trials	
Quasi-experimental method	
SAMPLE SIZE CRITERIA SELECTED	
LIFETIME ENERGY AND DEMAND SAVINGS FACTORS TO BE EVALUATED	
Degradation	
Rebound	
Other	

CHECKLIST FOR NET SAVINGS DETERMINATION	
NET SAVINGS FACTORS TO BE EVALUATED	
Free riders	
Spillover effects	
Other market effects	
Other(s)	
NET SAVINGS CALCULATION APPROACH SELECTED	
Self-reporting surveys	
Enhanced self-reporting surveys	
Large-scale consumption data analysis approaches (randomized controlled trial methods and quasi-experimental methods)	
Stipulated net-to-gross ratio	
Cross-sectional studies	
Top-down evaluations (or macro-economic models)	
CHECKLIST FOR AVOIDED EMISSIONS CALCULATIONS	
ELECTRICITY EFFICIENCY SAVINGS—GRID-CONNECTED	
Load emission rates	
Regional marginal baseload emission rates (using capacity factors or equivalent)	
Regional historical hourly emission rates	
Energy scenario modeling	
NATURAL GAS, FUEL OIL, AND NON-GRID-CONNECTED ELECTRIC GENERATING UNITS	
Default emission factor	
Source testing	

GENERIC EVALUATION CONSIDERATIONS (EXAMPLES)	
OVERALL GOALS	
Does the evaluation address the key policy, regulatory, and oversight needs for evaluation information?	
Will the program's 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 non-energy 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 trade-offs? 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	
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 measurements and impacts (emissions) boundaries been properly set?	

GENERIC EVALUATION CONSIDERATIONS (EXAMPLES)	
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 addressed in the sample design?	

Chapter 8: Notes

[96] See discussion of theory-based evaluation in Section 8.3.1.

[97] *Cross-border trading* means that energy savings obtained in one jurisdiction cannot be used to meet energy efficiency requirements in another jurisdiction. Should a national energy efficiency resource portfolio be established at some point, then consistent national EM&V standards and deemed savings values/calculations may become required. For more information on EM&V requirements that may be associated with a national energy efficiency or clean-resource standard, see State and Local Energy Efficiency Action Network. (April 2011). *National Energy Efficiency Evaluation, Measurement, and Verification (EM&V) Standard: Scoping Study of Issues and Implementation Requirements.* Prepared by Schiller, S.R.; Goldman, C.A.; Galawish, E.; LBNL Environmental Energy Technologies Division. www1.eere.energy.gov/seeaction/pdfs/emvstandard_scopingstudy.pdf.

[98] Kushler, M.; Nowak, S.; Witte, P. (February 2012). *A National Survey of State Policies and Practices for the Evaluation of Ratepayer-Funded Energy Efficiency Programs*. American Council for an Energy-Efficient Economy (ACEEE). Report Number U122. www.aceee.org/research-report/u122.

[99] See the Efficiency Valuation Organization (EVO): www.evo-world.org.

[100] A very common interactive factor is the relationship between (1) reductions in heat generated by retrofitted lighting systems in a building as they consume less electricity and (2) the possible requirements for increased space heating and/or reduced space cooling.

[101] Lazar, J.; Baldwin, X. (August 2011). V*aluing the Contribution of Energy Efficiency to Avoided Marginal Line Losses and Reserve Requirements.* Regulatory Assistance Project (RAP). www.raponline.org/document/download/id/4537.

[102] As of now, the baseline study contractor is separate from statewide contractors hired by the PSC for other activities listed in this table.

[103] May be done in conjunction with Northeast Energy Efficiency Partnerships (NEEP) effort.

Appendix A: Glossary

This glossary defines and explains terms used in this guide and in the evaluation, measurement, and verification (EM&V) of energy efficiency programs. Included are terms commonly used in the design and implementation of EM&V, terms often found in EM&V reports, and terms associated with efficiency measures, programs, and program strategies. Some of the definitions in this glossary reflect usage in the specific context of energy efficiency program administrators and implementers with an oversight (regulatory) body; thus, they may not be applicable to non-regulatory contexts such as in commercial agreements between energy services companies and their clients.

References used for preparation of this glossary are given at the end of this appendix. Glossaries tend to be living documents. For example, the well-regarded *"EM&V Forum Glossary of Terms,"* a project of the Regional Evaluation, Measurement and Verification Forum is regularly updated as a result of reviews by forum members and outside experts.[104]

Accuracy: A concept that refers to the relationship between the true value of a variable and an estimate of the value. The term can also be used in reference to a model or a set of measured data, or to describe a measuring instrument's capability.

Additionality: A criterion applied to, for example, greenhouse gas (GHG) projects, stipulating that project-based GHG reductions should only be quantified if the project activity "would not have happened anyway;" in other words, that the project activity (or the same technologies or practices it employs) would not have been implemented in its baseline scenario and/or that project activity emissions are lower than baseline emissions.

Adjusted Savings: See claimed savings and evaluated savings.

Administrator: An entity selected by a regulatory or other government organization to contract for and administer an energy efficiency portfolio within a specific geographic region and/or market. Typical administrators are utilities selected by a public service commission or a nonprofit or state government agency, as determined by legislation.

Allowances: The basic tradable commodity within a pollutant emissions trading systems. Allowances grant their holder the right to emit a specific quantity of pollution once (e.g., one ton of CO_2 (eq)). The total quantity of allowances issued by regulators dictates the total quantity of emissions possible under the system. At the end of each compliance period, each regulated entity must surrender sufficient allowances to cover their pollutant emissions during that period. 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 www.epa.gov/capandtrade/index.html.

Annual Demand Savings: The maximum reduction in electric or gas demand associated with energy efficiency activities in a given year within a defined boundary. More specific definitions of demand and thus demand savings occur within specific regulatory jurisdictions or electric control areas.

SAVINGS

Savings, or more accurately stated "savings estimates," from energy-efficiency measures, projects, programs, and portfolios are reported at various times in the lifecycle of the efficiency activity and with varying degrees of certainty. The two major points when they are reported are prior to and after the implementation of the activity. Different jurisdictions have different names for savings reports, what they contain, and whether and what adjustments or evaluation activities take place between pre-implementation and post-implementation. For example, prior to implementation, savings can be called Ex Ante *Savings* or *Projected Savings*. After implementation, there are even more variations in names and content, with titles including *ex post savings, adjusted savings, evaluated savings,* and *tracking estimates*. Examples of differences in these "after implementation" reports can include whether the results have been corrected for data errors, whether they have been verified (e.g., confirmed actual number of installations versus planning projections), and/or whether a third-party evaluator has generated or reviewed the reported values. Savings can also be indicated as first-year, annual demand or energy savings, and/or lifetime energy or demand savings values. They also can be indicated as gross savings and/or net savings values. **In this glossary, several savings reporting classifications are defined. Ideally, these terms and methods should be applied; however, it is critical that whenever savings are reported, the basis for the values indicated be made clear, rather than just relying on these common classifications, since their definition (and titles)—as well as how values are determined for each classification—can vary from jurisdiction to jurisdiction.**

Annual Energy Savings: The reduction in electricity use (kilowatt-hours) or in fossil fuel use (in thermal units) associated with energy efficiency activities in a given year within a defined boundary.

ASHRAE Guideline 14: American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) Guideline 14, *2002 Measurement of Energy and Demand Savings* (www.ashrae.org).

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

Attribution: Ascribing or establishing a causal relationship between action(s) taken by an entity and an outcome. For efficiency program evaluation, this is associated with the difference between net and gross savings. For example, an impact evaluation indicates that 30% of the gross energy savings associated with a ceiling fan incentive program could be attributed to the ENERGY STAR® labeling program and not the incentive program.

Avoided Costs: The forecasted economic benefits of energy savings. In the context of energy efficiency, these are the costs that are avoided by the implementation of an energy efficiency activity. Such costs are used in benefit-cost analyses of energy efficiency activities. Because efficiency activity reduces the need for electric generation, these costs include those associated with the cost of electric generation, transmission, distribution, and reliability. Typically, costs associated with avoided energy and capacity are calculated. Other costs avoided by the efficiency activity can also be included, among them the value of avoided emissions not already embedded in the generation cost, impact of the demand reduction on the overall market price for electricity, and avoided fuel or water. For natural gas efficiency programs, avoided costs can include components of the production, transportation, storage, and service that are variable to the amount of natural gas delivered to customers.

Baseline: Conditions, including energy consumption and demand, which would have occurred without implementation of the subject energy efficiency activity. Baseline conditions are sometimes referred to as "business-as-usual" conditions and are used to calculate project- and program-related savings. Baselines can also include definition of non-energy metrics that are being evaluated, such as air emissions and jobs.

Baseline Adjustments: For measurement and verification analyses, factors that modify baseline energy or demand values to account for independent variable values (conditions) in the reporting period.

Baseline Data: The baseline conditions of the facilities, market segment, generating equipment, or other area of focus of the subject efficiency activity.

Baseline Efficiency: The energy use of the baseline equipment, process, or standard that is being replaced by a more efficient approach to providing the same energy service (a subset of baseline). It is used to determine the energy savings obtained by the more efficient approach.

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

Baseline Scenario: A hypothetical description of what would have most likely occurred without implementation of the subject project or program.

Behavior Programs: Energy efficiency programs that use strategies intended to affect consumer energy use behaviors in order to achieve energy and/or energy demand savings. Programs typically include outreach, education, competition, rewards, benchmarking, and/or feedback elements. Such programs may rely on changes to consumers' habitual behaviors (e.g., turning off lights) or "one-time" behaviors (e.g., changing thermostat settings). In addition, these programs may target purchasing behaviors (e.g., purchases of energy-efficient products or services), often in combination with other programs (e.g., rebate programs or direct install programs). These programs are also distinguished by normally being evaluated using large-scale data analysis approaches involving experimental or quasi-experimental methods, versus deemed savings or measurement and verification approaches.

Benchmarking: A process that compares the energy, emissions, and other resource-related conditions of a facility against industry best practices or other benchmarks such as average per square foot energy consumption of similar buildings in the same city.

Benefit-Cost Ratio: The mathematical relationship between the benefits and costs associated with the implementation of energy efficiency measures, programs, practices, or emission reductions. The benefits and costs are typically expressed in dollars.

Benefit-Cost Test: The methodology used to compare the benefits of an investment with the costs; also called *cost-effectiveness* test. Five key benefit-cost tests have, with minor updates, been used for more than 20 years as the principal approaches for energy efficiency program evaluation. These five cost-effectiveness tests are the *participant cost test*, the *program administrator cost test,* the *ratepayer impact measure test,* the *total resource cost test,* and the *societal cost test.*

Bias: The extent to which a measurement or a sampling or analytic method systematically underestimates or overestimates a value. Some examples of types of bias include engineering model bias; meter bias; sensor placement bias; inadequate or inappropriate estimates of what would have happened absent a program or measure installation; a sample that is unrepresentative of a population; and selection of other variables in an analysis that are too correlated with the savings variable (or each other) in explaining the dependent variable (such as consumption).

Billing Analysis: A term used to define either (1) a specific measurement and verification (M&V) approach used to estimate project savings or (2) any analytic methodology used to determine project or program energy savings based on the use of the energy consumption data contained in consumer billing data. It compares billing data from program participant(s) over a period of time before the energy-efficient measures are installed at customer site(s) to billing data for a comparable period of time afterward. If used to describe an M&V approach, it is equivalent to IPMVP Option C, Whole Building Analysis. If used to describe an evaluation approach, it is comparable to the large-scale data analysis approach.

Billing Data: Data obtained from the electric or gas meter that is used to bill the customer for energy used in a particular billing period. In an evaluation context, *billing data* also refers to the customer billing records over time. Those records are used to conduct analyses of energy use before and after implementation of energy efficiency measures.

Building Commissioning: See *commissioning*.

Building Energy Simulation Model: Computer models based on physical engineering principals and/or standards used to estimate energy use and/or savings. These models usually incorporate site-specific data on customers and physical systems, such as square footage, weather, surface orientations, elevations, space volumes, construction materials, equipment use, lighting, and building occupancy. Building simulation models can usually account for interactive effects between end uses (e.g., lighting and HVAC), part-load efficiencies, and changes in external and internal heat gains/losses. Examples of building simulation models include DOE-2, EnergyPlus, and Carrier HAP.

Calibration: In economic, planning, or engineering modeling, the process of adjusting the components of the model to reflect reality as closely as possible, in order to prepare for the model's use in future applications. The term also applies to the process whereby metering and measurement equipment is periodically adjusted to maintain industry measurement standards.

California Measurement Advisory Council (CALMAC): An informal committee comprised 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 (www.calmac.org).

Claimed Savings: Values reported by an implementer or administrator, using their own staff and/or an evaluation consulting firm, after the subject energy efficiency activities have been completed; also called *tracking estimates, reported savings,* or in some cases, e*x post savings* (although ex post usually applies to evaluated savings [see *evaluated savings*]).

As with projected savings estimates, these values may use results of prior evaluations and/or values in a technical reference manual. However, they may be adjusted from projected savings estimates by correcting for any known data errors and actual installation rates and may also be adjusted with revised values for factors such as per-unit savings values, operating hours, and savings persistence rates. Claimed savings can be indicated as first-year, annual, and/or lifetime energy or demand savings values, and can indicated as gross savings and/or net savings values.

Co-Benefits: The impacts of an energy efficiency program other than the direct purpose (i.e., energy and demand savings) for which it was designed. See *non-energy benefits*.

Coefficient of Variation (CV): The mean (average) of a sample, divided by its standard error.

Coincident Demand: The 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. The peak of interest should be specified (e.g., "demand coincident with the utility system peak"). The following are examples of peak demand:

- Demand coincident with utility system peak load

- Demand coincident with independent system operator/regional transmission organization summer or winter peak or according to performance hours defined by wholesale capacity markets
- Demand coincident with high electricity demand days.

Commissioning: Often abbreviated as "Cx," a systematic quality assurance process associated with new construction that spans the entire design and construction process, helping ensure that a new building's performance meets owner expectations; sometimes referred to as *building commissioning*. Commissioning ensures that the new building operates as the owner intended and that building staff are prepared to operate and maintain its systems and equipment.

Common Practice: The predominant technology(ies) implemented or practice(s) undertaken in a particular region or sector. Common practices can be used to define a baseline.

Comparison Group: See *control group*.

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

Confidence: An indication of how close, expressed as a probability, the true value of the quantity in question is within a specified distance to the estimate of the value. Confidence is the likelihood that the evaluation has determined the true value of a variable within a certain estimated range. For example, a program that produces an estimate of 2% energy savings, with a 95% confidence interval of (1%, 3%) means that there is a 95% probability that the true program energy savings is between 1% and 3%. A smaller confidence interval implies that the estimate is more precise (e.g., a 2% energy savings estimate with a confidence interval of [1.5%, 2.5%] is more precise than a 2% energy savings estimate with a confidence interval of [1%, 3%]). See *precision*.

Control 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 treatment (participant) group; also called comparison group. The comparison group is used to isolate program effects from other factors that affect energy use.

Cooling Degree Days: The cumulative number of degrees in a month or year by which the mean temperature is above a set temperature, usually 18.3°C/65°F. See *degree days*.

Correlation: For a set of observations, such as for participants in an energy efficiency program, the extent to which high values for one variable are associated with high values of another variable for the same participant. For example, facility size and energy consumption usually have a high positive correlation.

Cost-Benefit and Cost-Effectiveness Analysis: Analysis that compares the benefits associated with a program or measure's outputs or outcomes with the costs (resources expended) to produce them. Cost-benefit analysis is typically conducted to determine the relationship of the program's benefits and costs, as a ratio. Cost-effectiveness analysis is generally undertaken to compare one program or program approach to other approaches, or options for the use of funds, to determine the relationship among the options. The terms are often interchanged in evaluation discussions. See *benefit-cost test*.

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). See *benefit-cost ratio*.

Cumulative Energy Savings: The summation of energy savings (e.g., megawatt-hours, therms) from multiple projects or programs over a specified number of years, taking into account the time of measure installation in the first year, annual energy savings for subsequent years, and the average life of the installed measures.

Custom Program: An energy efficiency program intended to provide efficiency solutions to unique situations not amenable to common or prescriptive solutions. Each custom project is examined for its individual characteristics, savings opportunities, efficiency solutions, and often, customer incentives.

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

Deemed Savings Calculation: An agreed-to (stipulated) engineering algorithm(s) used to calculate the energy and/or demand savings associated with an installed energy efficiency measure. These calculations are developed from common practice that is widely considered acceptable for the subject measure and its specific application. It may include stipulated assumptions for one or more parameters in the algorithm, but typically it requires users to input data associated with the actual installed measure into the algorithm(s).

Deemed Savings Value: An estimate of energy or demand savings for a single unit of an installed energy efficiency measure that (1) has been developed from data sources and analytical methods that are widely considered acceptable for the measure and purpose, and (2) is applicable to the situation being evaluated. Individual parameters or calculation methods can also be deemed; also called *stipulated savings value.*

Degree Days: For any individual day, an indication of how far that day's average temperature departed from a fixed temperature, usually 18.3°C/65°F. Heating degree days, which measure heating energy demand, quantify how far the average temperature fell below 65°F. Similarly, cooling degree days, which measure cooling energy demand, quantify how far the temperature averaged above 65°F. In both cases, smaller values represent less energy demand; however, values below 0 are set equal to 0, because energy demand cannot be negative. Furthermore, because energy demand is cumulative, degree day totals for periods exceeding one day are simply the sum of each individual day's degree days total. Degree days are used in calculations of heating and cooling loads and in evaluation regression analyses to adjust for differences in heating and cooling requirements between baseline and project scenarios.

Demand: The time rate of energy flow. It is the requirement for energy consumption of energy source(s) by an energy using system at a given instant or averaged over any designated interval of time. Demand usually refers to the amount of electric energy used by a customer or piece of equipment at a specific time, expressed in kilowatts (kW equals kWh/h) but can also refer to natural gas use at a point in time, usually as Btu/hr, kBtu/hr, therms/day, or cubic feet per day (ccf/day).

Demand Response (DR): The reduction of consumer energy use at times of peak use in order to help system reliability, reflect market conditions and pricing, or support infrastructure optimization or deferral of additional infrastructure. Demand response programs may include contractually obligated or voluntary curtailment, direct load control, and pricing strategies.

Demand Savings: The reduction in electric or gas demand from the baseline to the demand associated with the higher-efficiency equipment or installation. This term is usually applied to billing demand to calculate cost savings or peak demand for equipment sizing purposes.

Demand-Side Management (DSM): Strategies used to manage v energy demand, including energy efficiency, load management, fuel substitution, and load building.

Dependent Variable: Term used in regression analysis or other analyses seeking to explain the relationship among variables to quantify the variable that is being explained by the other (independent) variables.

Direct Emissions: Emissions from sources within an entity's organizational boundaries that are owned or controlled by the entity, including stationary combustion emissions, mobile combustion emissions, process emissions, and fugitive emissions. 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).

Direct Install Program: An energy efficiency program design strategy involving the direct installation of measures in customer premises by a contractor sponsored by the program. Such programs generally involve one-for-one replacement of existing equipment with more efficient equipment, and may include a customer rebate.

Discount Rate: An interest rate applied to a stream of future costs and/or monetized benefits to convert those values to a common period, typically the current or near-term year, to measure and reflect the time value of money. It is used in benefit-cost analysis to determine the economic merits of proceeding with the proposed project, and in cost-effectiveness analysis to compare the value of projects. The discount rate for any analysis is either a nominal discount rate or a real discount rate. A nominal discount rate is used in analytic situations when the values are in then-current or nominal dollars (reflecting anticipated inflation rates). A real discount rate is used when the future values are in constant dollars, and can be approximated by subtracting expected inflation from a nominal discount rate.

Diversity: That characteristic of a variety of electric loads whereby individual maximum demands of each load usually occur at different times.

Diversity Factor: The ratio of the sum of the demands of a group of users to their coincident maximum demand during a specified period of time (e.g., summer or winter).

Education Programs: Programs primarily intended to educate customers about energy efficient technologies or behaviors or provide information about programs that offer energy efficiency or load-reduction information or services. These programs may provide indirect energy or demand savings.

Effective Useful Life (EUL): An estimate of the duration of savings from a measure. EUL is estimated through various means, including median number of years that the energy efficiency measures installed under a program are still in place and operable. Also, EUL is sometimes defined as the date at which 50% of Installed units are still in place and operational. See *measure life*.

Efficiency: The ratio of the useful energy delivered by a dynamic system (e.g., a machine, engine, or motor) to the energy supplied to it over the same period or cycle of operation. The ratio is usually determined under specific test conditions.

Emission Factor: A representative value that relates the quantity of a pollutant released to the atmosphere with an activity associated with the release of that pollutant. These factors are usually expressed as the weight of pollutant divided by a unit weight, volume, distance, or duration of the activity emitting the pollutant (e.g., pounds of a pollutant per million Btu of heat input or pounds of emissions per MWh of electricity produced). Such factors facilitate estimation of emissions from various sources of air pollution. They are based on available data of acceptable quality, and are generally assumed to be representative of averages for that pollutant.

End Use: General categories of energy efficiency measures reflecting the type of services provided (e.g., lighting, HVAC, motors, and refrigeration).

End-Use Metering: The direct measuring of energy consumption or demand by specific end-use equipment, typically as part of load research studies or to measure the impacts of demand-side management programs.

Energy Conservation: Term used to reflect doing with less of a service in order to save energy. The term is often unintentionally used instead of energy efficiency.

Energy Efficiency: The use of less energy to provide the same or an improved level of service to the energy consumer; or, the use of less energy to perform the same function.

Energy Efficiency Activity: Any of a wide range of actions that are anticipated to result in the more efficient use of energy. Energy efficiency measures are a subset of energy efficiency activities, as energy efficiency activities include non-technology specific actions such as education.

Energy Efficiency Measure: At an end-use energy consumer facility, an installed piece of equipment or system; a strategy intended to affect consumer energy use behaviors; or modification of equipment, systems, or operations that reduces the amount of energy that would otherwise have been used to deliver an equivalent or improved level of end-use service.

Energy Savings: Reduction in electricity use in kilowatt-hours or in fossil fuel use in thermal unit(s).

ENERGY STAR®: A joint program of the U.S. Environmental Protection Agency and the U.S. Department of Energy designed to reduce energy use and the impact on the environment. The ENERGY STAR label is awarded to products that meet applicable energy efficiency guidelines and to homes and commercial buildings that meet specified energy efficiency standards. The program provides a range of energy management tools, primarily computer-based, for businesses.

Engineering Methods: The use of standard formulas or models based on those formulas, typically accepted by ASHRAE, as the basis for calculating energy use.

Engineering Model: Engineering equations used to calculate energy use 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 use or savings as a function of measurable attributes of customers, facilities, or equipment (e.g., lighting use = watts x hours of use).

Equipment Life: The number of years that a measure is installed and operates until failure.

Error: The deviation of measurements from the true value of the variable being observed; also called *measurement error*.

Evaluated Savings: Savings estimates reported by an independent, third-party evaluator after the subject energy efficiency activities have been implemented and an impact evaluation has been completed; also called ex post, or more appropriately, *ex post evaluated savings*.

The designation of “independent” and “third party” is determined by those entities involved in the use of the evaluations and may include evaluators retained, for example, by the administrator or a regulator. These values may rely on claimed savings for factors such as installation rates and a technical reference manual for values such as per-unit savings values and operating hours. These saving estimates may also include adjustments to claimed savings or projected savings for data errors, per-unit savings values, operating hours, installation rates, savings persistence rates, or other considerations. Evaluated savings can be indicated as first-year, annual, and/or lifetime energy or demand savings values. They also can be indicated as gross savings and/or net savings values.

Evaluation: The conduct of any of a wide range of assessment studies and other activities aimed at determining the effects of a program and understanding or documenting program performance, program or program-related markets and market operations, program-induced changes in energy efficiency markets, levels of demand or energy savings, or program cost-effectiveness. Market assessment, monitoring and evaluation, and measurement and verification are aspects of evaluation.

Evaluator: A person or entity that conducts evaluations.

Ex Ante Savings: See *projected savings*.

Experimental Design: A method of estimating the impact of a program or other event (e.g., a medication, a procedure), in which outcomes between at least two randomly assigned groups are compared.

Ex Post Savings: See *claimed savings* and *evaluated savings*.

Ex Post Evaluated Savings: See *evaluated savings*.

External Validity: The condition in which an impact estimate that is internally valid for a given program population and time frame can be generalized and applied to new situations (e.g., new populations, future years).

FEMP M&V Guidelines: U.S. Department of Energy Federal Energy Management Program's *2008 M&V Guidelines: Measurement and Verification for Federal Energy Projects*.

Free Driver, Non-Participant: A program non-participant who has adopted particular energy efficiency measure(s) or practice(s) as a result of the evaluated program. See *spillover*.

Free Driver, Participant: A program participant who has adopted additional or incremental energy efficiency measure(s) or practice(s) as a result of the evaluated program, but which were not directly induced by the program. See *spillover*.

Free Rider: A program participant who would have implemented the program's measure(s) or practice(s) in the absence of the program. Free riders can be (1) total, in which the participant's activity would have completely replicated the program measure; (2) partial, in which the participant's activity would have partially replicated the program measure; or (3) deferred, in which the participant's activity would have partially or completely replicated the program measure, but at a future time beyond the program's time frame.

Fuel Switching: Using an alternative fuel (usually of lower carbon intensity) to produce required energy.

Gross Market Savings: The change in energy consumption and/or demand that results from energy efficiency programs, codes and standards, and naturally occurring adoption, which have a long-lasting savings effect. Gross market savings generally do not include temporary reductions in energy use from changes in weather, income, energy prices, and other structural economic changes such as in industry composition.

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

Heating Degree Days: The cumulative number of degrees in a month or year by which the mean temperature falls below a fixed temperature, usually 18.3°C/65°F. See *degree days*.

Home Energy Rating System (HERS): An indexing system, associated with ENERGY STAR®, used in residential new construction to rate the pre- and post-construction of new homes to highlight and indicate the degree of energy efficiency embedded in the construction. The HERS Index is a scoring system established by the Residential Energy Services Network (RESNET) in which a home built to the specifications of the HERS Reference Home (based on the 2006 International Energy Conservation Code) scores a HERS Index of 100, while a net zero energy home scores a HERS Index of 0. The lower a home's HERS Index, the more energy efficient it is in comparison to the HERS Reference Home. Each 1-point decrease in the HERS Index corresponds to a 1% reduction in energy consumption compared to the HERS Reference Home.

HVAC: Heating, ventilation, and air conditioning.

Impact Evaluation: An evaluation of the program-specific, directly or indirectly induced changes (e.g., changes in energy and/or demand use) associated with an energy efficiency program.

Implementer: An entity selected and contracted with or qualified by a program administrator to provide products and services to consumers either directly or indirectly.

Incentive: A financial strategy intended to encourage a change in behavior related to energy use. Incentives can take various forms. Customer incentives are commonly used in energy efficiency programs as rebates for individual measures or as buy-downs in more custom-oriented projects. Performance or shareholder incentives are monies that are established in a planning period to encourage program administrators to attain specified levels of savings during the program year.

Incremental Annual Savings: The difference between the amount of energy savings acquired or planned to be acquired as a result of energy efficiency activities in one year, and the amount of energy savings acquired or planned to be acquired as a result of the energy efficiency activities in the prior year.

Incremental Cost: The difference between the cost of existing or baseline equipment or service and the cost of alternative energy-efficient equipment or service.

Independent Variables: The explanatory factors (e.g., weather or occupancy) in a regression model that are assumed to affect the variable under study (e.g., energy use).

Indirect Emissions: Emissions that are a consequence of activities that take place within the organizational boundaries of an entity, but occur at sources owned or controlled by another entity. For example, emissions of electricity used by a manufacturing entity that occur at a power plant represent the manufacturer's indirect emissions. Indirect emissions are typically the source of avoided emissions for electric energy efficiency measures.

Indirect Energy (Demand) Savings (Indirect Program Energy Savings): The use of the words "indirect savings" or "indirect program savings" refers to programs that are typically information, education, marketing, or outreach programs in which the program's actions are expected to result in energy savings achieved through the actions of the customers exposed to the program's efforts, without direct enrollment in an program that has energy-savings goals.

Inspections: Site visits to facilities treated under an energy efficiency program that document the existence, characteristics, and operation of baseline or project equipment and systems, as well as factors that affect energy use. Inspections may or may not include review of commissioning or retro-commissioning documentation.

Installation Rate: The percentage of measures that are incented by an energy efficiency program that are actually installed in a defined period of time. The installation rate is calculated by dividing the number of measures installed by the number of measures incented by an energy efficiency program in a defined period of time.

Interactive Effects: The influence of one technology's application on the energy required to operate another application. An example is the reduced heat in a facility as a result of replacing incandescent lights with CFLs, and the resulting need to increase space heating from another source, usually oil- or gas-fired. With respect to IPMVP Options A and B, interactive effects in energy use or demand occurring beyond the measurement boundary of the M&V analysis.

Internal Validity: Refers to how well an evaluation was conducted (e.g., design, how variables were measured, what was/wasn't measured) and how confidently one can conclude that the observed effect(s) were produced solely by the independent variable and not extraneous ones. For impact evaluations, this is related to whether the savings impacts are valid for the specific program being evaluated, the given program participant population, and the given time frame of the evaluation. This is often compared to *external validity*.

International Performance Measurement and Verification Protocol (IPMVP): A guidance document with a framework and definitions describing the four M&V approaches; a product of the Efficiency Valuation Organization (www.evo-world.org).

Leakage: In its broadest terms, the concept that the effect of an activity or outcome expected to occur and remain within a defined boundary flows outside the boundary, leading to unintended results. In efficiency programs, an example of leakage is when a measure is incented by a program (with the associated costs and assumed savings) but is installed outside of the program's jurisdiction. In the context of air regulation, such as a cap-and-trade program, an example of leakage is a shift of electricity generation from sources subject to the cap-and-trade program to higher-emitting sources not subject to the program. 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.

Levelized Cost: The result of a computational approach used to compare the cost of different projects or technologies. The stream of each project's net costs is discounted to a single year using a discount rate (creating a net present value) and divided by the project's expected lifetime output (megawatt-hours or therms).

Lifetime Cost Per Kilowatt-hour or Therm: The cost associated with a piece of equipment (supply-side or demand-side), energy efficiency program, or total portfolio during its expected life in relation to (divided by) the electricity or gas that it produces or saves over its lifetime. The annual costs are usually discounted back to a single year using an appropriate discount rate.

Lifetime Demand Savings: The expected demand savings over the lifetime of an installed measure(s), project(s), or program(s). It may be calculated by multiplying the annual peak demand reduction associated with a subject measure(s) by the expected useful lifetime of the measure(s). It may include consideration of technical degradation and possibly the rebound effect. Savings can be gross or net. For electricity, it can be expressed in units of kilowatt-years.

Lifetime Energy Savings: The expected energy savings over the lifetime of an installed measure(s), project(s), or program(s). It may be calculated by multiplying the annual energy use reduction associated with a subject measure(s) by the expected useful lifetime of the subject measure(s). It may include consideration of technical degradation and possibly the rebound effect. Savings can be gross or net.

Load Factor: A percentage indicating the difference between the amount of electricity or natural gas a consumer used during a given time span and the amount that would have been used if the use had stayed at the consumer's highest demand level during the whole time. The term also means the percentage of capacity of an energy facility, such as a power plant or gas pipeline, that is used in a given period of time. It is also the ratio of the average load to the peak load during a specified time interval.

Load Management: Steps taken to reduce power demand at peak load times or to shift some of it to off-peak times. Load management may coincide with peak hours, peak days, or peak seasons. Load management may be pursued by persuading consumers to modify behavior or by using equipment that regulates some electric consumption. This may lead to complete elimination of electric use during the period of interest (load shedding) and/or to an increase in electric demand in the off-peak hours as a result of shifting electric use to that period (load shifting).

Load Shapes: Representations such as graphs, tables, and databases that show the time-of-use pattern of customer or equipment energy use. These are typically shown over a 24-hour or whole-year (8,760 hours) period.

Logic Model: The graphical representation of a program theory showing the connection among activities, their outputs, and subsequent short-term, intermediate, and long-term outcomes. Often, the logic model is displayed with these elements in boxes, and the causal flow is shown by arrows from one to the others in the program logic. It can also be displayed as a table, with the linear relationship presented by the rows in the table.

Main Meter: The meter that measures the energy used for the whole facility. There is at least one meter for each energy source and possibly more than one per source for large facilities. Typically, utility meters are used, but dataloggers may also be used as long as they isolate the load for the facility being studied. When more than one meter per energy source exists for a facility, the main meter may be considered the accumulation of all the meters involved.

Marginal Cost: The sum that has to be paid for the next increment of product or service. The marginal cost of electricity is the price to be paid for kilowatt-hours above and beyond those supplied by presently available generating capacity.

Market: The commercial activity (e.g., manufacturing, distributing, buying, and selling) associated with products and services that affect energy use.

Market Assessment: An analysis that provides an assessment of how and how well a specific market or market segment is functioning with respect to the definition of well-functioning markets or with respect to other specific policy objectives. A market assessment generally includes a characterization or description of the specific market or market segments, including a description of the types and number of buyers and sellers in the market, the key actors that influence the market, the type and number of transactions that occur on an annual basis, and the extent to which market participants consider energy efficiency an important part of these transactions. This analysis may also include an assessment of whether a market has been sufficiently transformed to justify a reduction or elimination of specific program interventions. Market assessment can be blended with strategic planning analysis to produce recommended program designs or budgets. One particular kind of market assessment effort is a baseline study, or the characterization of a market before the commencement of a specific intervention in the market for the purpose of guiding the intervention and/or assessing its effectiveness later.

Market Barrier: Any characteristic of the market for an energy-related product, service, or practice that helps to explain the gap between the actual level of investment in, or practice of, energy efficiency and an increased level that would appear to be cost beneficial to the consumer.

Market Effect: A change in the structure of a market or the behavior of participants in a market that is reflective of an increase (or decrease) in the adoption of energy efficient products, services, or practices and is causally related to market interventions (e.g., programs). Examples of market effects include increased levels of awareness of energy efficient technologies among customers and suppliers, increased availability of energy efficient technologies through retail channels, reduced prices for energy efficient models, build out of energy efficient model lines, and—the end goal—increased market share for energy efficient goods, services, and design practices.

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 Event: The broader circumstances under which a customer considers adopting an energy efficiency product, service, or practice. Types of market events include, but are not necessarily limited to (1) new construction, or the construction of a new building or facility; (2) renovation, or the updating of an existing building or facility; (3) remodeling, or a change in an existing building; (4) replacement, or the replacement of equipment, either as a result of an emergency such as equipment failure or as part of a broader planned event; and (5) retrofit, or the early replacement of equipment or refitting of a building or facility while equipment is still functioning, often as a result of an intervention into energy efficiency markets.

Market Participants: The individuals and organizations participating in transactions with one another within an energy efficiency market or markets, including customers and market actors.

Market Penetration Rate: A measure of the diffusion of a technology, product, or practice in a defined market, as represented by the percentage of annual sales for a product or practice, the percentage of the existing installed stock for a product or category of products, or the percentage of existing installed stock that uses a practice.

Market Saturation: A percentage indicating the proportion of a specified end-user market that contains a particular product. An example would be the percentage of all households in a given geographical area that have a certain appliance. Studies conducted to obtain this information within the residential sector are referred to as *residential appliance saturation studies (RASS)*.

Market Sectors: General types of markets that a program may target or in which a service offering may be placed. Market sectors include categories such as agricultural, commercial, industrial, government, and institutional.

Market Segments: A part of a market sector that can be grouped together as a result of a characteristic similar to the group. For example, within the residential sector are market segments such as renters, owners, multifamily, and single-family.

Market Theory: A theoretical description of how a market operates relative to a specific program or set of programs designed to influence that market. Market theories typically include the identification of key market actors, information flows, and product flows through the market, relative to a program designed to change the way the market operates. Market theories are typically grounded upon the information provided from a market assessment but can also be based on other information. Market theories often describe how a program intervention can take advantage of the structure and function of a market to transform the market. Market theories can also describe the key barriers and benefits associated with a market and describe how a program can exploit the benefits and overcome the barriers.

Market Transformation: A reduction in market barriers resulting from a market intervention, as evidenced by a set of market effects that is likely to last after the intervention has been withdrawn, reduced, or changed.

Measure: [verb] Use of an instrument to assess a physical quantity or use of a computer simulation to estimate a physical quantity.

Measure: [noun] See energy *efficiency measure*.

Measure Life: The length of time that a measure is expected to be functional; sometimes referred to as *expected useful life*. Measure life is a function of equipment life and measure persistence. *Equipment life* is the number of years that a measure is installed and will operate until failure. *Measure persistence* takes into account business turnover, early retirement of installed equipment, and other reasons measures might be removed or discontinued.

Measurement and Verification (M&V): A subset of program impact evaluation that is associated with the documentation of energy savings at individual sites or projects using one or more methods that can involve measurements, engineering calculations, statistical analyses, and/or computer simulation modeling. M&V approaches are defined in the IPMVP.

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

Measure Penetration: The fraction of annual market sales captured by a more efficient measure or system at a given point in time; also called *market penetration*. For example, the market penetration of CFLs was 20% in the Florida market in 2005.

Measure Persistence: The duration of an energy-consuming measure, taking into account business turnover, early retirement of installed equipment, technical degradation factors, and other reasons measures might be removed or discontinued.

Measure Retention Study: An assessment of (1) the length of time the measure(s) installed during the program year are maintained in operating condition, and (2) the extent to which there has been a significant reduction in the effectiveness of the measure(s).

Measure Saturation: The fraction of a total market (e.g., consumers, buildings) captured by a specific efficiency activity at a given point in time; also called *market saturation*. For example, in 2005, 20% of the commercial buildings that are more than 100,000 square feet in Portland had high-efficiency chillers that exceeded minimum efficiency standards.

Metered Data: Data collected over time through a meter for a specific energy using, end-use system (e.g., lighting and HVAC), or location (e.g., floors of a building or a whole premise). Metered data may be collected over a variety of time intervals. Metered data usually refers to electricity or gas data.

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: The collection of relevant measurement data over time at a facility, including but not limited to energy consumption or emissions data (e.g., energy and water consumption, temperature, humidity, volume of emissions, hours of operation) for the purpose of savings analysis or to evaluate equipment or system performance.

Naturally Occurring Efficiency: The effects of energy-related decisions, by both program participants and non-participants, that would have been made in the absence of the program; alternatively, the expected average efficiency of one or more measures or systems in the absence of all publicly funded energy efficiency programs. It can be part of a baseline determination.

Net Present Value (NPV): The value of a stream of cash flows converted to a single sum in a specific year, usually the first year of the analysis. It can also be thought of as the equivalent worth of all cash flows relative to a base point called the present.

Net Savings: The change in energy consumption and/or demand that is attributable to a particular energy efficiency program. This change in energy use and/or demand may include, implicitly or explicitly, consideration of factors such as free ridership, participant and non-participant spillover, and induced market effects. These factors may be considered in how a baseline is defined (e.g., common practice) and/or in adjustments to gross savings values.

Net-to-Gross (NTG) Ratio: 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. The factor itself may be made up of a variety of factors that create differences between gross and net savings, commonly including free riders and spillover. Can be applied separately to either energy or demand savings.

New Construction: Residential and nonresidential buildings that have been newly built or have added major additions.

Nominal: For dollars, "nominal" means the figure representing the actual number of dollars exchanged in each year, without accounting for the effect of inflation on the value or purchasing power. For interest or discount rates, "nominal" means that the rate includes the rate of inflation (the real rate plus the inflation rate approximately equals the nominal rate).

Non-Energy Effects or Non-Energy Benefits (NEB): The identifiable non-energy impacts associated with program implementation or participation; also referred to as *non-energy impacts (NEI)* or *co-benefits*. Examples of NEBs include avoided emissions and other environmental benefits, productivity improvements, jobs created, reduced program administrator debt and disconnects, and higher comfort and convenience level of the participant. The value is most often positive, but may also be negative (e.g., the cost of additional maintenance associated with a sophisticated, energy-efficient control system).

Non-Participant: Any consumer who was eligible but did not participate in the subject efficiency program, in a given program year.

Normalized Annual Consumption (NAC) Analysis: A regression-based method that analyzes monthly energy consumption data and adjusts the consumption data to eliminate annual or other periodic fluctuations in an influencing factor (e.g., weather on heating and cooling needs) based on a historical normal or average pattern of the influencing factor.

Offset: Program mechanism that allows an entity to neutralize a portion or all of its regulated, capped emissions contribution by orchestrating or funding projects that are not subject to regulation (i.e., an emissions cap).

Panel Data Model: An estimation analysis model that contains many data points over time rather than averaged, summed, or otherwise aggregated data.

Participant: A consumer that received a service offered through the subject efficiency program, in a given program year; also called *program participant*. The term "service" is used in this definition to suggest that the service can be a wide variety of inducements, 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.

Participant Cost Test (PCT): A cost-effectiveness test that measures the economic impact to the participant of adopting an energy efficiency measure.

Peak Demand: The maximum level of hourly demand during a specified period. The peak periods most commonly identified are annual and seasonal (summer and winter).

Peak Load: The highest electrical demand within a particular period of time. Daily electric peaks on weekdays typically occur in late afternoon and early evening. Annual peaks typically occur on hot summer days.

Persistence: See s*avings persistence rate* and *measure persistence*.

Persistence Study: A study to assess changes in program impacts over time. See *savings persistence rate* and *measure persistence*.

Portfolio: Either (1) 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 (2) 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, Achievable: The amount of energy or demand savings within a defined geographical area or population that can be achieved in response to specific energy efficiency program designs, delivery approaches, program funding, and measure incentive levels; sometimes referred to as *market potential, program potential,* or *realistic potential*.

Potential, Economic: Refers to the subset of the technical potential that is economically cost-effective as compared to conventional options.

Potential Studies: Studies conducted to assess market baselines and future savings that may be expected for different technologies and customer markets over a specified time horizon.

Potential, Technical: An estimate of energy savings based on the assumption that all existing equipment or measures will be replaced with the most efficient equipment or measure that is both available and technically feasible over a defined time horizon, without regard to cost or market acceptance.

Precision: The indication of the closeness of agreement among repeated measurements of the same physical quantity. Precision is a measure of how statistically confident evaluators can be that the estimated impact of a program is close to the true impact of a program. An estimate with a smaller confidence interval is said to be more precise. It is also used to represent the degree to which an estimated result in social science (e.g., energy savings) would be replicated with repeated studies.

Prescriptive Program: An energy efficiency program focused on measures that are one-for-one replacements of the existing equipment and for which fixed customer incentives can be developed based on the anticipated similar savings that will accrue from their installation.

Primary Effects: Effects that the project or program are intended to achieve. For efficiency programs, this is primarily a reduction in energy use (and/or demand) 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: An activity, strategy, or course of action undertaken by an implementer or administrator. Each program is defined by a unique combination of the program strategy, market segment, marketing approach, and energy efficiency measure(s) included. Programs consist of a group of projects with similar characteristics and installed in similar applications.

Program Administrator Cost Test (PACT): See *utility/program administrator cost test*.

Program Incentive: An incentive, generally monetary, that is offered to a customer through an energy efficiency program to encourage the customer to participate in the program. The incentive is intended to overcome one or more barriers that keep the customer from taking the energy efficiency activity on his own.

Program Theory: A presentation of the goals of a program, incorporated with a detailed presentation of the activities that the program will use to accomplish those goals and the identification of the causal relationships between the activities and the program's effects. Program theory is often the basis for the logic model.

Program Year (PY): The calendar year approved for program implementation. Note that program years can be shorter than 12 months if programs are initiated mid-year.

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

Projected Savings: Values reported by an implementer or administrator before the subject energy efficiency activities are implemented; also called *planning estimates* or *ex ante estimates*.

Projected savings are typically estimates of savings prepared for program and/or portfolio design or planning purposes. These values are typically based on pre-program or portfolio estimates of factors such as per-unit savings values, operating hours, installation rates, and savings persistence rates. These values may use results of prior evaluations and/or values in a technical reference manual. They can be indicated as first-year, annual, and/or lifetime energy or demand savings values. They can be indicated as gross savings and/or net savings values.

Random Assignment: A condition where each household or entity in the study population is randomly assigned to either the control group or the treatment group based on a random probability, as opposed to being assigned to one group or the other based on some characteristic of the household (e.g., location, energy use, or willingness to sign up for the program). Randomization creates a control group that is statistically identical to the treatment group, in both observable and unobservable characteristics, such that any difference in outcomes between the two groups can be attributed to the treatment with a high degree of confidence.

Randomized Controlled Trial (RCT): A type of experimental program evaluation design in which energy consumers in a given population are randomly assigned into two groups: a treatment group and a control group. The outcomes for these two groups are compared, resulting in program energy savings estimates.

Ratepayer Impact Measure (RIM) Test: A cost-effectiveness test that measures the impact on utility operating margin and whether rates would have to increase to maintain the current levels of margin if a customer installed energy efficient measures. The test measures what happens to customer bills or rates due to changes in utility revenues and operating costs caused by the program.

Real: For dollars, "real" means that the dollars are expressed in a specific base year in order to provide a consistent means of comparison after accounting for inflation. For interest and discount rates, "real" means the inflation rate is not included (the nominal rate minus the inflation rate approximately equals the real rate).

Realization Rate: Used in several contexts for comparing one savings estimate with another. The primary and most meaningful application is the ratio of evaluated gross savings to claimed gross savings (versus comparing net and gross savings estimates, which are best defined with a net-to-gross ratio). Basis for the ratio not being 1.0 can include several considerations, such as (1) adjustments for data errors, (2) differences in implemented measure counts as a result of verification activities, and/or (3) other differences revealed through the evaluation process, such as changes in baseline assumptions.

Rebate: See *incentive*.

Rebate Program: An energy efficiency program in which the program administrator offers a financial incentive for the installation of energy-efficient equipment.

Rebound Effect: A change in energy-using behavior that yields an increased level of service accompanied by an increase in energy use that occurs as a result of taking an energy efficiency action; also called take back. The result of this effect is that the absolute savings associated with the direct energy efficiency action is impacted (usually reduced) by the resulting behavioral change. The rebound effect can be considered as a factor in determining savings persistence rate.

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.

Regression Model: A mathematical model based on statistical analysis where the dependent variable is quantified based on its relationship to the independent variables that are said to determine its value. In so doing, the relationship between the variables is estimated statistically from the data used.

Reliability: The quality of a measurement process that would produce similar results on (1) repeated observations of the same condition or event, or (2) multiple observations of the same condition or event by different observers. 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.

Representative Sample: A sample that has approximately the same distribution of characteristics as the population from which it was drawn. Such samples may be randomly selected or not. Random samples selected from the same population as participants are more representative of the program participants.

Resource Acquisition Program: A program designed to achieve directly energy and/or demand savings; also called *retrofit program*. Such a program generally involves encouraging customers to replace existing equipment with more-efficient equipment.

Retro-commissioning: Often abbreviated as "RCx," a systematic method for investigating how and why an existing building's systems are operated and maintained, and for identifying ways to improve overall building performance. Retro-commissioning improves a building's operations and maintenance (O&M) procedures to enhance overall building performance.

Retrofit: Energy efficiency activities undertaken in existing residential or nonresidential buildings, where existing inefficient equipment or systems are replaced by energy-efficient equipment or systems or where efficient equipment or systems are added to an existing facility (e.g., addition of thermal insulation).

Retrofit Isolation: The savings measurement approach defined in IPMVP Options A and B, as well as 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. IPMVP Option A involves "Key Parameter Measurement" and IPMVP Option B involves "All Parameter Measurement."

Retrofit Program: An energy efficiency program that provides incentives, information, and technical support to encourage customers to replace existing and operating equipment or systems with more energy-efficient equipment or systems that provides the same function, or to add energy efficient equipment or systems to an existing facility (e.g., addition of thermal insulation).

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 reliable.

Sample: In program evaluation, a portion of the population selected to represent the whole. Differing evaluation approaches rely on simple or stratified (based on some characteristic of the population) samples.

Sample Design: The approach used to select the sample units.

Sampling Error: An error that arises because the data are collected from a part, rather than the whole of the population. It is usually measurable from the sample data in the case of probability sampling.

Savings Persistence Rate: Percentage of first-year energy or demand savings expected to persist over the life of the installed energy efficiency equipment. It is developed by conducting surveys of installed equipment several years after installation to determine presence and operational capability of the equipment.

Seasonal Energy Efficiency Ratio (SEER): The total cooling output of a central air-conditioning unit (in Btu) during its normal use period for cooling divided by the total electrical energy input (in watt-hours) during the same period, as determined using specified federal test procedures.

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). Secondary effects can be positive or negative.

Snap Back: See *rebound effect*.

Societal Cost Test (SCT): A cost-effectiveness test that measures the net economic benefit to the utility service territory, state, or region, as measured by the total resource cost test, plus indirect benefits such as environmental benefits.

Spillover (Participant and Non-Participant): Reductions in energy consumption and/or demand caused by the presence of an energy efficiency program, beyond the program-related gross savings of the participants and without direct financial or technical assistance from the program. There can be participant and/or non-participant spillover. *Participant spillover* is the additional energy savings that occur as a result of the program's influence when a program participant independently installs incremental energy efficiency measures or applies energy-saving practices after having participated in the energy efficiency program. *Non-participant spillover* refers to energy savings that occur when a program non-participant installs energy efficiency measures or applies energy savings practices as a result of a program's influence.

Spillover Rate: Estimate of energy savings attributable to spillover effects expressed as a percent of savings installed by participants through an energy efficiency program.

Statistically Adjusted Engineering (SAE) Models: A category of statistical analysis models that incorporate the engineering estimate of savings as a dependent variable. The regression coefficient in these models is the percentage of the engineering estimate of savings observed in changes in energy use. For example, if the coefficient on the SAE term is 0.8, this means that the customers are on average realizing 80% of the savings from their engineering estimates.

Stipulated Values: See *deemed savings*.

Stratified Random Sampling: A sampling method where the population is divided into X units of subpopulations, called strata, that are non-overlapping and together make up the entire population. A simple random sample is taken of each strata to create a sample based upon stratified random sampling.

Stratified Ratio Estimation: A sampling method that combines a stratified sample design with a ratio estimator to reduce the coefficient of variation by using the correlation of a known measure for the unit (e.g., expected energy savings) to stratify the population and allocate samples from strata for optimal sampling. Stratified ratio estimation can reduce the number of sites, observations, and thus evaluation costs required to achieve a given level of precision by using the correlation of a known measure for the unit (e.g., expected energy savings) to stratify the population and allocate sample requirements for each strata.

Structured Interview: An interview in which the questions to be asked, their sequence, and the detailed information to be gathered are all predetermined. These are used where maximum consistency across interviews and interviewees is needed.

Take Back Effect: See *rebound effect*.

Technical Degradation Factor: A multiplier used to account for time- and use-related change in the energy savings of a high-efficiency measure or practice relative to a standard-efficiency measure or practice due to technical operational characteristics of the measures, including operating conditions and product design.

Technical Reference Manual (TRM): A resource document that includes information used in program planning and reporting of energy efficiency programs. It can include savings values for measures, engineering algorithms to calculate savings, impact factors to be applied to calculated savings (e.g., net-to-gross ratio values), source documentation, specified assumptions, and other relevant material to support the calculation of measure and program savings—and the application of such values and algorithms in appropriate applications.

Total Resource Cost (TRC) Test: A cost-effectiveness test that measures the net direct economic impact to the utility service territory, state, or region.

Tracking Estimate: See *claimed savings*.

Treatment Group: The group of consumers that receive the subject program's intervention or "treatment" (i.e., the program participants).

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.

Upstream Program: A program that provides information and/or financial assistance to entities in the delivery chain of high-efficiency products at the retail, wholesale, or manufacturing level. Such a program is intended to yield lower retail prices for the products.

Utility/Program Administrator Cost Test: A cost-effectiveness test that measures the change in the amount the utility must collect from the customers every year to meet an earnings target (e.g., a change in revenue requirement); also called *program administrator cost test (PACT)* and also known as the *utility cost test*. In several states, this test is referred to as the p*rogram administrator cost test*. In those cases, the definition of the "utility" is expanded to program administrators (utility or third-party).

Verification: An independent assessment that the program has been implemented per the program design. For example, the objectives of measure installation verification are to confirm (1) the installation rate, (2) that the installation meets reasonable quality standards, and (3) that the measures are operating correctly and have the potential to generate the predicted savings. Verification activities are generally conducted during on-site surveys of a sample of projects. Project site inspections, participant phone and mail surveys, and/or implementer and consumer documentation review are typical activities associated with verification. Verification may include one-time or multiple activities over the estimated life of the measures. It may include review of commissioning or retro-commissioning documentation. Verification is a subset of evaluation and, as such, can also include review and confirmation of evaluation methods used, samples drawn, and calculations used to estimate program savings.

Whole-Building Calibrated Simulation Approach: A savings measurement approach (defined in IPMVP Option D and ASHRAE Guideline 14) that involves the use of an approved computer simulation program to develop a physical model of the building in order to determine energy and demand savings. The simulation program is used to model the energy used by the facility before and after the retrofit. The pre- or post-retrofit models are calibrated with measured energy use and demand data as well as weather data.

Whole-Building Metered Approach: A savings measurement approach (defined in the IPMVP Option C and ASHRAE Guideline 14) that determines energy and demand savings through the use of whole-facility energy (end-use) data, which may be measured by utility meters or data loggers. This approach may involve the use of monthly utility billing data or data gathered more frequently from a main meter.

Workforce Education and Training Programs: Programs primarily intended for building maintenance engineers, HVAC contractors, engineers, architects, maintenance personnel, and others. These programs provide information about energy efficiency concepts, recommended energy-efficient technologies or behaviors, and/or programs that offer energy efficiency or load-reduction information, products or services. These programs may provide indirect energy or demand savings.

Glossary References

California Commissioning Collaborative. (2006). California Commissioning Guide: New Buildings and California Commissioning Guide: Existing Buildings. www.cacx.org/resources/commissioning-guides.html.

California Public Utilities Commission. (April 2006). *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals. Prepared by The TecMarket Works Team. www.calmac.org/events/EvaluatorsProtocols_Final_AdoptedviaRuling_06-19-2006.pdf.*

Efficiency Valuation Organization. (September 2010). International Performance Measurement and Verification Protocol. EVO 10000—1:2010. Concepts and Options for Determining Energy and Water Savings Volume 1. www.evo-world.org.

Eto, J.; Prahl, R.; Schegel, J. (1996). A Scoping Study on Energy-Efficiency Market Transformation by California Utility DSM Programs. Lawrence Berkeley National Laboratory. http://eetd.lbl.gov/ea/ems/reports/39058.pdf.

National Action Plan for Energy Efficiency. (2007). Guide for Conducting Energy Efficiency Potential Studies. Optimal Energy, Inc. www.epa.gov/cleanenergy/documents/suca/potential_guide.pdf.

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

National Action Plan for Energy Efficiency. (2008). Understanding Cost-Effectiveness of Energy Efficiency Programs: Best Practices, Technical Methods, and Emerging Issues for Policy-Makers. Prepared by Energy and Environmental Economics, Inc. (E3) and Regulatory Assistance Project (RAP). www.epa.gov/cleanenergy/documents/suca/cost-effectiveness.pdf.

National Action Plan for Energy Efficiency. (2009). Energy Efficiency as a Low-Cost Resource for Achieving Carbon Emissions Reductions. ICF International, Inc. www.epa.gov/cleanenergy/documents/suca/ee_and_carbon.pdf.

Northeast Energy Efficiency Partnerships. (July 2011). EM&V Forum GLOSSARY OF TERMS Version 2.1. PAH Associates. A project of the Regional Evaluation, Measurement and Verification Forum. http://neep.org/uploads/EMV%20Forum/EMV%20Products/EMV_Glossary_Version_2.1.pdf.

State and Local Energy Efficiency Action Network. (May 2012). Evaluation, Measurement, and Verification (EM&V) of Residential Behavior-Based Energy Efficiency Programs: Issues and Recommendations. Prepared by Todd, A.; Stuart, E.; Schiller, S.; Goldman, C.; Lawrence Berkeley National Laboratory. www1.eere.energy.gov/seeaction/pdfs/emv_behaviorbased_eeprograms.pdf.

World Resources Institute and World Business Council for Sustainable Development. (November 2005). The GHG Protocol for Project Accounting. www.ghgprotocol.org/files/ghgp/ghg_project_protocol.pdf.

Appendix B: Other Evaluation Categories and Approaches

This appendix provides a brief introduction to process and market effects evaluations, cost-effectiveness analysis, and impact evaluations using "top-down" approaches. The material in this appendix is intended to supplement the other sections of the guide, which are focused on "bottom-up" impact evaluations.

B.1 PROCESS, MARKET EFFECTS, AND COST-EFFECTIVENESS EVALUATIONS

The following subsections introduce three non-impact types of evaluations: process, market, and cost-effectiveness. However, because cost-effectiveness analyses rely on the documentation of program impacts, these analyses are often considered a component of impact evaluations, and program cost-effectiveness indicators are thus often included in impact evaluation reports. Table B.1 compares these three evaluation types, plus impact evaluations.

B.1.1 Process Evaluations

The goal of process evaluations is to produce better and more cost-effective programs. Process evaluations meet this goal by assessing the processes a program undergoes during implementation, documenting program goals and objectives from a variety of perspectives, and describing program strengths and weaknesses so that success is highlighted and improvements can be made in a timely manner. Thus, process evaluations examine the efficiency and effectiveness of program implementation procedures and systems. Typical process evaluation results involve recommendations for changing a program's structure, implementation approaches, and goals.

These evaluations usually consist of asking questions of those involved in the program, analyzing their answers, and comparing results to established best practices. Whereas it is typically required that an independent third-party evaluator is involved in conducting

TABLE B.1: Program Evaluation Types

EVALUATION TYPE	DESCRIPTION	EXAMPLES USES
Impact Evaluation	Quantifies direct and indirect changes associated with the subject program(s)	Determines the amount of energy and demand saved
Process Evaluation	Indicates how the procedures associated with program design and implementation are performing from both the administrator's and the participants' perspectives	Identifies how program designs and processes can be improved
Market Effects Evaluation	Analyzes how the overall supply chain and market for energy efficiency products have been affected by the program	Characterizes changes that have occurred in efficiency markets and whether they are attributable to and sustainable with or without the program
Cost-Effectiveness Evaluation	Quantifies the costs of program implementation and compares them with program benefits	Determines whether an energy efficiency program is a cost-effective investment compared with other programs and energy supply resources

impact evaluations, for process evaluations, jurisdictions might recommend (but not require) them to be conducted by independent third-party evaluators; however, the use of third-party process evaluators is a best practice. Use of a trusted party for process evaluation is important for successful process evaluation so that the evaluator can gather the necessary data and provide feedback in a manner that is productive (e.g., not considered threatening by the recipient of the feedback).

Process evaluations are particularly valuable in the following situations:

- Benefits are higher/lower than expected and/or are being achieved more quickly/slowly than expected.
- There is limited program participation or stakeholders are slow to begin participating.
- The program is a greater success than anticipated.
- The program has a slow start-up.
- Participants are reporting problems.
- The program appears not to be cost effective.
- The program is built around a new concept that could be replicable for other populations, technologies, etc.

As part of a process evaluation, a logic model may be developed for the program (or possibly a set of logic models for a complete portfolio of programs). A program's theory and logic model serve as a roadmap to guide the systematic approach of a process evaluation. A program logic model is a visual representation of the program's theory that illustrates a set of interrelated program activities that combine to produce a variety of outputs that lead to key outcomes (see sidebar in Chapter 7 on Theory-Based Evaluation: A Guiding Principle for MT Evaluation). Logic models can be linked to performance indicators that provide ongoing feedback to program managers. The models usually flow top to bottom and are often organized according to five basic categories:

- **Program inputs:** financial, staffing, and infrastructure resources that support the activity
- **Program activities:** overarching activities that describe what the program is doing (e.g., marketing and rebate processing)
- **Outputs:** metrics resulting from the activities, and that tend to be measurable "bean counting" results (e.g., provide outreach events at five community fairs)
- **Short- to intermediate-term outcomes:** expected outcomes resulting from program activities, with goals attached to those outcomes when possible. (e.g., target energy savings and recruitment into the program)
- **Long-term outcomes and goals:** ideal, sustainable outcomes resulting from program activities (e.g., "all eligible customers participate in the program" and "increase customer awareness of program offerings").

These logic model categories indicate the intended and expected results of activities. Expected short-, medium-, and long-term outcomes tend to define program goals at a high level and also specify market effects (i.e., expected program outcomes). In this manner, process evaluation is part of a continuum linking impact and market effects evaluations.

TABLE B.2: Elements of Typical Process Evaluations

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

The primary mechanism of process evaluations is data collection (e.g., surveys, questionnaires, and interviews) from administrators, designers, participants (e.g., facility operators, business owners, renters, or homeowners), implementation staff (including contractors, subcontractors, and field staff), trade allies (e.g., mechanical contractors, architects, and engineers) and key policymakers. 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 collection and analysis of relevant data from third-party sources (e.g., equipment vendors or retailers). Process evaluations can be operated continuously, perhaps as part of a continuous improvement effort, or at intervals (e.g., as a new program is being implemented, whenever there are major changes in a program, in response to issues noted in first set of bullets above, and/or just every two to three years).

Table B.2 lists examples of program elements typically assessed during a process evaluation.

B.1.2 Market Effects Evaluations

The goal of market effects evaluations is to characterize and quantify the effects of a program on supplier promotion and customer adoption of the targeted energy efficiency measures, regardless of whether those suppliers and customers participated in the program. Effects that cannot be captured by program records are particularly important for certain kinds of initiatives, including "upstream" promotions of mass-market goods, such as light bulbs and consumer electronics as well as training programs aimed at inducing engineers and contractors to adopt energy efficiency design and specification practices. Studies have shown that even straightforward equipment rebate programs may have effects "outside the program" by exposing contractors and large customers to the benefits of efficient technologies. This in turn leads to increased specification of efficient technologies on projects that do not receive program support. In some cases, market effects evaluation results can be combined with impact evaluation findings to estimate program-induced energy savings that were not tracked by the program itself.

Other market studies include potential studies (see sidebar) and market baseline studies. Potential studies investigate how much saving may be available through various measures and baseline studies look at indicators of market development before the program intervention.

Market effects studies are usually associated with programs that have a specific market transformation focus. There are many definitions of market transformation, although it is often considered the ultimate goal of publicly and consumer-funded energy efficiency programs. In this guide, the definition of *market transformation* is: a reduction in market barriers resulting from a market intervention, as evidenced by a set of market effects, that is likely to last after the intervention has been withdrawn, reduced, or changed.

Market effects evaluations often involve a significant undertaking, because they require collection and analysis of data from a wide range of market actors, as well as analysis of those data against a background developed out of secondary sources. Market effects are sometimes called the ultimate test of a program's success, answering the question: "Will energy efficiency (best) practices continue in the marketplace, even after the current program ends?" The difference between a market change and a market effect is attribution: the ability to trace back a change in the market to a specific program or group of programs. The following is a definition of market effects from a well-referenced 1996 study:[105]

> **Market effect:** a change in the structure of a market or the behavior of market participants that is reflective of an increase in the adoption of energy-efficient products, services, or practices and is causally related to market intervention(s) (e.g., programs). Examples of market effects include increased levels of awareness of energy-efficient technologies among customers and suppliers, increased availability of efficient technologies through retail channels, reduced prices for efficient models, build-out of efficient model lines, and—the end goal—increased market share for efficient goods, services, and design practices.

POTENTIAL STUDIES

Another form of market study (although not formally an "evaluation") is called a potential study. Potential studies are conducted before a program is implemented in order to assess market baselines and future savings potentials for different efficiency technologies, strategies, or approaches in different 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), and 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). Findings also help managers identify the program's key markets and clients and how to best serve the intended customers.

Examples of the questions that a market effects evaluation might answer are as follows:

- **Are the entities that undertook energy efficiency projects undertaking additional projects or incorporating additional technologies in their facilities that were not directly induced by the program?** This might indicate that 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 energy efficiency concepts.
- **Are manufacturers, distributors, retailers, vendors, and others involved in the supply chain of energy efficiency products (and services) changing their product offerings—for example, how are they marketing them, pricing them, stocking them?** The answers can indicate how the supply chain is adapting to changes in supply of and demand for efficiency products.

Structuring a market effects evaluation entails consideration of several levels or stages, with the ultimate goal generally understood to be the increased adoption of energy efficiency goods and services in the general market leading to energy savings. Energy savings are the ultimate goal of programs seeking to cause market effects (i.e., the intended long-term outcome). The following list suggests a hierarchy of precursors to that goal:

- **Early Acceptance:** proliferation of models and manufacturers, purchase by frontrunners and enthusiasts
- **Take-off Phase:** customer awareness, visibility on shelves and in inventories, perceptible levels of market share in the supply channels
- **Maturity:** all major competitors offer energy efficient models; codes and standards include energy efficient models
- **Energy Savings:** energy savings attributable to the program are associated with acceleration of these developments.

In general, the achievement of goals at each of the higher levels of the hierarchy requires accomplishments at the lower levels. As a result, tracking goals at each stage not only provides feedback on performance with respect to that goal itself, but also provides evidence that effects at the next-higher levels can be attributed to the program.

Goals will typically be set and tracked for different time frames and for different purposes. While energy savings are the ultimate market effects goal, in most cases, savings cannot be measured meaningfully without several years of information; even then, they will usually not have the same level of accuracy as impact evaluations of direct resource acquisition savings. To credit measure adoption and associated savings to a program, it must be shown that the increased energy efficiency adoption, the longer-term market effects, and the participant effects have all occurred essentially in the manner and in the order specified by the program theory. And this, for most programs, takes a number of years to reach this point.

In 2009, a comprehensive white paper study on market transformation and market effects was prepared.[106] Table B.3, from a presentation by the principle author of this white paper, indicates approaches for assessing market effects, including attribution.

As can be deduced from the above discussion and Table B.3, the market effects evaluation can easily overlap with the spillover analyses conducted as part of an impact evaluation. In fact, many of the techniques used to quantify market effects can be applied in estimating spillover savings.

B.1.3 Cost-Effectiveness Analyses[107, 108]

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

In 1983, California's *Standard Practice for Cost-Benefit Analysis of Conservation and Load Management Programs* manual (SPM) developed five cost-effectiveness tests for evaluating energy efficiency programs. These approaches, with minor updates, continue to be used today and are the principal approaches used for evaluating energy efficiency programs across the United States.[109]

The five tests vary in terms of (1) perspectives (project participants, ratepayers, utilities, or society), (2) their applicability to different program types, (3) the cost and benefit elements included in the calculation, (4) the methods by which the cost and benefit elements are computed, and (5) the uses of the results. Most regulated utility energy 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 as follows on page 144:

- **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

TABLE B.3: Approaches for Assessing Market Effects, Including Attribution

BASIC SOURCE/RELATIVE ADVANTAGES	LIMITATIONS
SURVEYS OF CUSTOMER PURCHASES • Can be deployed quickly, relatively inexpensively, and repeatedly over extended time frames • Can be deployed in program and non-program areas • Generally produces reliable data on number of purchases/adoptions	• Limited accuracy on key details: exact number, timing, efficiency rating of purchases • Non-response bias a problem, particularly in early stages of market development • Difficult to validate results in absence of some comparison to sales or program volumes
SURVEYS OF SUPPLY-SIDE ACTORS • Taps into close knowledge of local markets • Respondents sufficiently knowledgeable to provide accurate information on product features	• Difficult to build measures of sales volume—may need to be content with estimates of market share • In many jurisdictions, population available to be sampled is small • Difficult to validate results in absence of some comparison to sales or program volumes
SHIPMENT AND SALES DATA • Conceptually, the most accurate and detailed measure of adoption: quantity, efficiency, timing	• Requires negotiated cooperation of manufacturers and retailers; risk of dropouts • Difficult to obtain coverage of all sectors, time periods, and regions (and may be costly) • Quality control is difficult
CUSTOMER-REPORTED FREE RIDERSHIP AND SPILLOVER • Can be deployed quickly, relatively inexpensively, and repeatedly over extended time frames • Can probe adoption process and decisions • Consistent with current Performance Earnings Basis (PEB) methods now in force in California	• For non-participants, requires that customers be aware of the program and able to judge its impact on adoption decisions
CROSS-SECTIONAL METHODS • Closest to conventional social science research methods; intuitively satisfying • Data provide insight into exogenous factors, working of market beyond program boundary	• Increasingly difficult to find non-program areas • Difficult to verify comparability of non-program areas • Appears to be effective only in time-limited periods • Logistically demanding and time consuming
EXPERT JUDGING • Focuses insights from experienced market participants and observers • Results can be expressed in terms of net adoptions • In some cases, can be deployed fairly rapidly	• Not a statistical estimation process • Difficult to identify and account for factors affecting individual judgments

Source: Rosenberg, M. (June 2010) "Market Effects and Market Transformation: Their Role in Program Design and Evaluation." EPA EM&V Webinar. www.emvwebinar.org.

both the participants' and the utility's costs. It combines the perspectives of participants and non-participants, which is why it is also often called an "all ratepayers" perspective. 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 lifecycle basis, considering savings and costs that accrue over the lifetime of installed energy efficiency equipment or systems. This is a commonly applied cost-effectiveness test.

- **Program Administrator Cost Test (PACT).** The PACT measures the net costs of a demand-side management program as a resource option based on the costs incurred by the program administrator (often 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). The PACT is also a commonly applied test.
- **Participant Cost Test (PCT).** The PCT assesses cost effectiveness from the participating consumer's perspective by calculating the quantifiable benefits and costs to the consumer of participating in a program. Because many consumers do not base their decision to participate entirely on quantifiable criteria, this test is not necessarily a complete measure of all the benefits and costs a participant perceives.
- **Societal Cost Test (SCT).** The SCT, 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 lifecycle costs and benefits, the societal test accounts for externalities (e.g., environmental benefits), excludes tax credit benefits, and uses a (often lower) societal discount rate.
- **Ratepayer Impact Measure (RIM) Test.** The RIM test only applies to utility programs. It examines the potential impact that the energy efficiency program has on rates overall. The net benefits are the avoided cost of energy (same as PACT). The net costs include the overhead and incentive costs (same as PACT) but also include utility lost revenues from customer bill savings. Historically, reliance on the RIM test has limited energy efficiency investment, as it is the most restrictive of the five cost-effectiveness tests.

The basic structure of each cost-effectiveness test involves a calculation of the total benefits and the total costs in dollar terms from a certain vantage point to determine whether or not the overall benefits exceed the costs. A test is positive if the benefit-to-cost ratio is greater than one, and negative if it is less than one—with, of course, proper consideration of uncertainties in the inputs used in the calculation. Results are reported either in net present value (NPV) dollars (method by difference) or as a ratio (i.e., benefits/costs). Table B.4 outlines the basic approach underlying cost-effectiveness tests.

Each of the tests provides a different kind of information about the impacts of energy efficiency programs from different vantage points in the energy system. On its own, each test provides a single stakeholder perspective. Together, multiple tests provide a comprehensive approach to answering key questions: "Is the program effective overall?" "Is it balanced?" "Are some costs or incentives too high or too low?" "What is the effect on rates?" "What adjustments are needed to improve the alignment?"

Overall, the results of all five cost-effectiveness tests provide a more complete picture than the use of any one test alone. The TRC and SCT cost tests help to answer whether energy efficiency is cost-effective overall. The PCT, PACT, and RIM help to answer whether the selection of measures and design of the program is balanced from participant, utility, and non-participant perspectives, respectively. Looking at the cost-effectiveness tests together helps to characterize the attributes of a program or measure to enable decision making, to determine whether some measures or programs are too costly, whether some costs or incentives are too high or too low, and what adjustments need to be made to improve distribution of costs and benefits among stakeholders. The scope of the benefit and cost components included in each test is summarized in Table B.5 and Table B.6.

The broad categories of costs and benefits included in each cost-effectiveness test are consistent across all regions of the country and applications. However, the specific components included in each test may vary across different regions, market structures, and utility types. For example, transmission and distribution investment may be considered deferrable through energy efficiency in some areas and not in others. Likewise, the TRC and SCT may consider just natural gas or electricity resource savings in some cases, but also include co-benefits of other savings streams (such as water and fuel oil) in others.

Also, for the SCT, how the "non-monetized benefits" in Tables B.5 and B.6 are determined is an evolving area. In particular, benefits that in the past could not be monetized (e.g., air quality impacts) now can be assigned monetary values and, in fact, need to be assigned such values in order to be used in cost-effectiveness equations. Also, non-energy benefits, which in the past might have been ignored, are being shown to have significant value. These include economic development and employment benefits as more money is spent on local services and products because of the efficiency investments.

TABLE B.4: The Five Principal Cost-Effectiveness Tests Used in Energy Efficiency

TEST	ACRONYM	KEY QUESTION ANSWERED	SUMMARY OF APPROACH
Participant cost test	PCT	Will the participants benefit over the measure life?	Comparison of costs and benefits of the customer installing the measure
Program administrator cost test	PACT	Will utility bills increase?	Comparison of program administrator costs to supply-side resource costs
Ratepayer impact measure	RIM	Will utility rates increase?	Comparison of administrator costs and utility bill reductions to supply-side resource costs
Total resource cost test	TRC	Will the total costs of energy in the utility service territory decrease?	Comparison of program administrator and customer costs to utility resource savings
Societal cost test	SCT	Is the utility, state, or nation better off as a whole?	Comparison of society's costs of energy efficiency to resource savings and non-cash costs and benefits

California Public Utilities Commission (CPUC). (2001). California Standard Practice Manual: Economic Analysis of Demand-Side Programs and Projects. www.energy.ca.gov/greenbuilding/documents/background/07-J_CPUC_STANDARD_PRACTICE_MANUAL.PDF.

National Action Plan for Energy Efficiency. (2008). Understanding Cost-Effectiveness of Energy Efficiency Programs: Best Practices, Technical Methods, and Emerging Issues for Policy-Makers. Prepared by Energy and Environmental Economics, Inc. (E3) and Regulatory Assistance Project (RAP). www.epa.gov/cleanenergy/documents/suca/cost-effectiveness.pdf.

TABLE B.5: Description of Benefits and Costs Included in Each Cost-Effectiveness Test

<table>
<tr><th>TEST</th><th>BENEFITS</th><th>COSTS</th></tr>
<tr><td rowspan="2">PCT</td><td colspan="2">Benefits and costs from the perspective of the customer installing the measure</td></tr>
<tr><td>Incentive payments
Bill savings
Applicable tax credits or incentives</td><td>Incremental equipment costs
Incremental installation costs</td></tr>
<tr><td rowspan="2">PACT</td><td colspan="2">Perspective of utility, government agency, or third party implementing the program</td></tr>
<tr><td>Energy-related costs avoided by the utility
Capacity-related costs avoided by the utility, including generation, transmission, and distribution</td><td>Program overhead cost
Utility/program administrator incentive costs
Utility/program administrator installation costs</td></tr>
<tr><td rowspan="2">RIM</td><td colspan="2">Impact of efficiency measure on non-participating ratepayers overall</td></tr>
<tr><td>Energy-related costs avoided by the utility
Capacity-related costs avoided by the utility, including generation, transmission, and distribution</td><td>Program overhead cost
Utility/program administrator incentive costs
Utility/program administrator installation costs
Lost revenue due to reduced energy bills</td></tr>
<tr><td rowspan="2">TRC</td><td colspan="2">Benefits and costs from the perspective of all utility customers (participants and non-participants) in the utility service territory</td></tr>
<tr><td>Energy-related costs avoided by the utility
Capacity-related costs avoided by the utility, including generation, transmission, and distribution
Additional resource savings (i.e., gas and water if utility is electric)
Monetized environmental and non-energy benefits (see Section 4.9)
Applicable tax credits (see Section 6.4)</td><td>Program overhead costs
Program installation costs
Incremental measure costs (weather paid by the customer or utility)</td></tr>
<tr><td rowspan="2">SCT</td><td colspan="2">Benefits and costs to all the utility service territory, state, or nation as a whole</td></tr>
<tr><td>Energy-related costs avoided by the utility
Capacity-related costs avoided by the utility, including generation, transmission, and distribution
Additional resource savings (i.e., gas and water if utility is electric)
Non-monetized benefits (and costs) such as cleaner air or health impacts</td><td>Program overhead costs
Program installation costs
Incremental measure costs (weather paid by the customer or utility)</td></tr>
</table>

California Public Utilities Commission (CPUC). (2001). California Standard Practice Manual: Economic Analysis of Demand-Side Programs and Projects. www.energy.ca.gov/greenbuilding/documents/background/07-J_CPUC_STANDARD_PRACTICE_MANUAL.PDF.

National Action Plan for Energy Efficiency. (2008). Understanding Cost-Effectiveness of Energy Efficiency Programs: Best Practices, Technical Methods, and Emerging Issues for Policy-Makers. Prepared by Energy and Environmental Economics, Inc. (E3) and Regulatory Assistance Project (RAP). www.epa.gov/cleanenergy/documents/suca/cost-effectiveness.pdf.

TABLE B.6: Summary of Benefits and Costs Included in Each Cost-Effectiveness Test

COMPONENT	PCT	PACT	RIM	TRC	SCT
Energy and capacity-related avoided costs		Benefit	Benefit	Benefit	Benefit
Additional resource savings				Benefit	Benefit
Non-monetized benefits					Benefit
Incremental equipment and installation costs	Cost			Cost	
Program overhead costs		Cost	Cost	Cost	Cost
Incentive payments	Benefit	Cost	Cost		
Bill savings	Benefit		Cost		

California Public Utilities Commission (CPUC). (2001). California Standard Practice Manual: Economic Analysis of Demand-Side Programs and Projects. www.energy.ca.gov/greenbuilding/documents/background/07-J_CPUC_STANDARD_PRACTICE_MANUAL.PDF.

National Action Plan for Energy Efficiency. (2008). Understanding Cost-Effectiveness of Energy Efficiency Programs: Best Practices, Technical Methods, and Emerging Issues for Policy-Makers. Prepared by Energy and Environmental Economics, Inc. (E3) and Regulatory Assistance Project (RAP). www.epa.gov/cleanenergy/documents/suca/cost-effectiveness.pdf.

B.2 TOP-DOWN IMPACT EVALUATION

Top-down impact evaluation refers to methods that rely on aggregate energy consumption data or per-unit energy consumption indices (e.g., energy consumption per-unit of output or per capita) defined by sector, utility service territory, state, region, or country as the starting point for determining energy savings. Top-down evaluation focuses on the bottom line—reductions in energy use (and/or demand) for a state, region, or utility service territory. This gives top-down evaluation a direct link to (1) demand forecasting and resource planning, and (2) emissions accounting and forecasting, as used for greenhouse gas mitigation goals.

Figure B.1 compares the top-down with the bottom-up impact evaluation approaches that are discussed in the body of this guide. At present, virtually all energy efficiency program evaluations conducted in the United States rely on bottom-up approaches.

Top-down approaches start from aggregate data, such as state-level data for energy consumption, and then attempt to correlate any changes in energy consumption with measures of energy efficiency actions, such as expenditures or savings, using macro-economic models. The main advantages of top-down evaluation methods over bottom-up methods are their potentially lower evaluation costs due to relatively modest data requirements and the potential for directly estimating net program savings at the sector, state, regional, and national levels. The primary potential drawbacks of top-down evaluation are the difficulty in attributing energy consumption changes to specific energy efficiency policies and/or particular programs and actions.

A metric that can be considered the output of top-down evaluation is gross market savings. These are the energy savings resulting from energy efficiency programs, codes and standards, and naturally occurring adoption, and which have a long-lasting savings effect. Such gross market savings sometimes do not include temporary reductions in energy use from changes in weather, income, energy prices, and other structural economic changes, such as in industry composition. Figure B.2 shows a graphical illustration of the concept behind estimating gross market savings.[110]

During the last two decades, many energy efficiency practitioners and policymakers have expressed growing interest in the use of top-down methods for documenting the system-wide impacts and gross market savings of energy efficiency initiatives. Interest in top-down methods has grown from policymakers' and evaluation researchers' concerns that bottom-up evaluations have not properly accounted for effects of free ridership, spillover, and energy efficiency measure interactions—particularly in large program portfolios and in situations where energy consumer funds are used, such as in utility-sponsored efficiency programs. Top-down evaluations should also be less expensive to implement than bottom-up evaluations. Thus, research on top-down evaluation has been directed largely toward estimation of energy consumer-funded energy efficiency program savings, both the gross market energy savings and the portion attributable to the programs being evaluated.

Top-down energy efficiency evaluation methods are generally less developed than bottom-up methods in the energy efficiency field. However, at time of the publication of this guide, two macro-consumption (top-down) pilot studies are under way. Sponsored by the California Public Utilities Commission (CPUC), these studies are attempts to test the effectiveness and reliability of different top-down evaluation approaches and to determine whether they can be applied consistently to project market gross savings and to attribute savings to utility-sponsored energy efficiency investments. The results of these pilots can be found at the CALMAC website (www.calmac.org) and the CPUC website (www.cpuc.ca.gov/PUC/energy/Energy+Efficiency).

FIGURE B.1: Comparison of bottom-up versus top-down evaluation

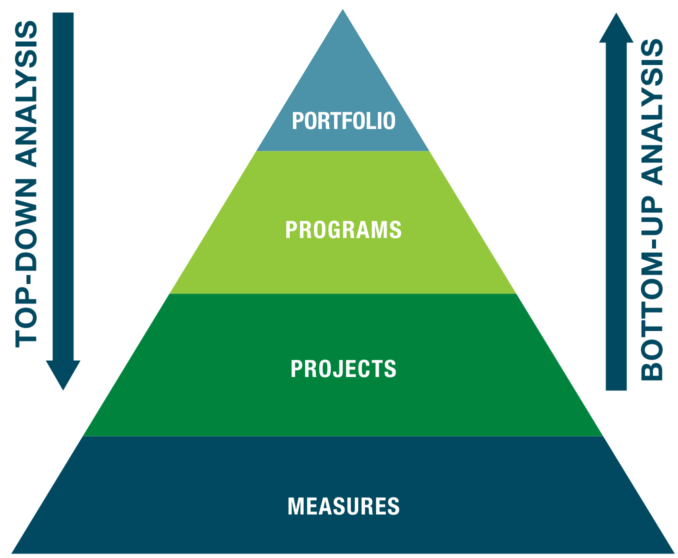
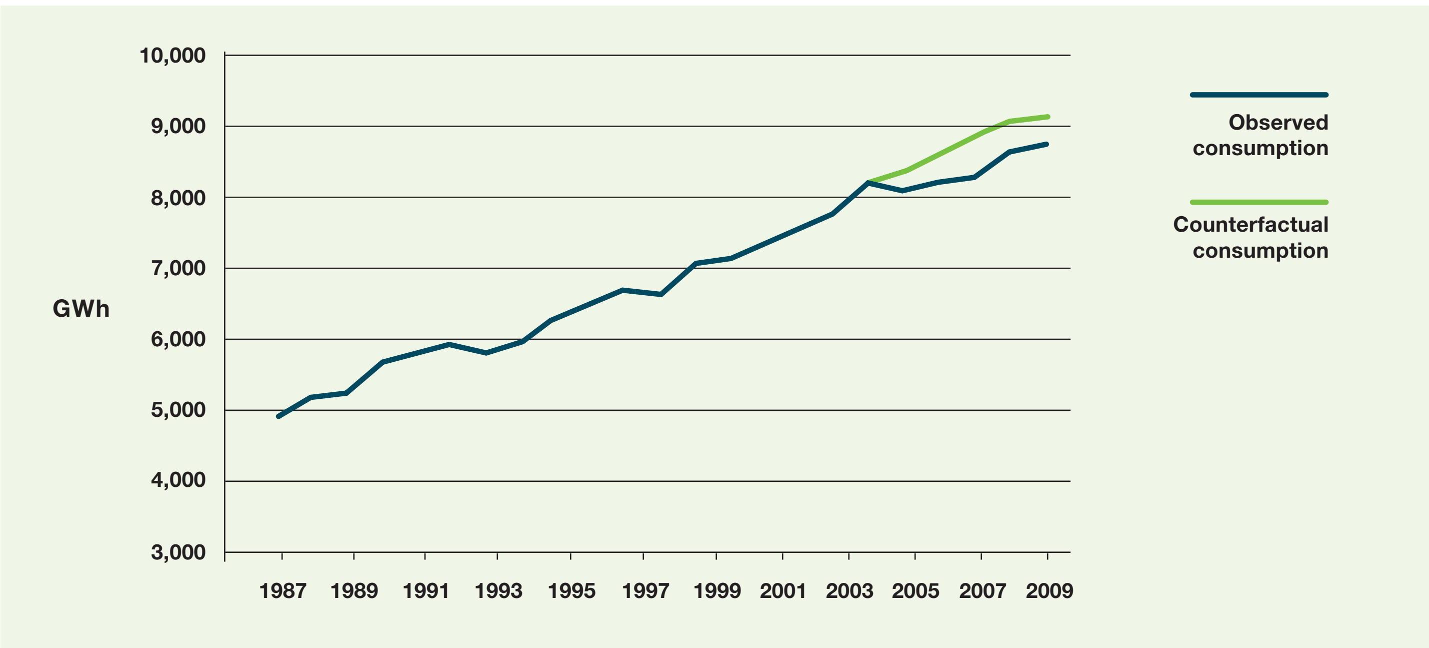


FIGURE B.2: Graphical illustration of estimation of market gross savings



Source: Stewart, J.; Haeri, M.H. (July 2011). Critical Review and Recommendations on Top-Down Evaluation. White paper. The Cadmus Group, Inc. Prepared for the California Public Utilities Commission (CPUC).

Appendix B: Notes

[105] Eto, J.; Prahl, R.; Schegel, J. (1996). *A Scoping Study on Energy-Efficiency Market Transformation by California Utility DSM Programs.* Lawrence Berkeley National Laboratory. http://eetd.lbl.gov/ea/ems/reports/39058.pdf.

[106] Rosenberg, M.; Hoefgen, L. (March 2009). *Market Effects and Market Transformation: Their Role in Energy Efficiency Program Design and Evaluation*. Prepared for California Institute for Energy and Environment (CIEE) by KEMA, Inc.; Nexus Market Research. http://uc-ciee.org/downloads/mrkt_effts_wp.pdf.

[107] Much of this subsection is taken (in some cases word for word) from the National Action Plan for Energy Efficiency. (2008). *Understanding Cost-Effectiveness of Energy Efficiency Programs: Best Practices, Technical Methods, and Emerging Issues for Policy-Makers*. Prepared by Energy and Environmental Economics, Inc. (E3) and Regulatory Assistance Project (RAP). www.epa.gov/cleanenergy/documents/suca/cost-effectiveness.pdf.

[108] Another recent summary report of cost-effectiveness tests is: Woolf, T.; Malone, E.; Takahashi, K.; Steinhurst, W. (July 23, 2012). *Best Practices in Energy Efficiency Program Screening: How to Ensure that the Value of Energy Efficiency is Properly Accounted For. Synapse Energy Economics, Inc.* Prepared for National Home Performance Council. www.synapse-energy.com/Downloads/SynapseReport.2012-07.NHPC.EE-Program-Screening.12-040.pdf.

[109] California Public Utilities Commission (CPUC). (2001). *California Standard Practice Manual: Economic Analysis of Demand-Side Programs and Projects.* www.energy.ca.gov/greenbuilding/documents/background/07-J_CPUC_STANDARD_PRACTICE_MANUAL.PDF. See also www.cpuc.ca.gov/PUC/energy/Energy+Efficiency/EM+and+V for the 2007 SPM Clarification Memo.

[110] Stewart, J.; Haeri, M.H. (July 2011). C*ritical Review and Recommendations on Top-Down Evaluation.* White paper. The Cadmus Group, Inc. Prepared for the California Public Utilities Commission (CPUC).

Appendix C: Resources

This appendix provides a listing of references that provide a body of knowledge developed over the last several decades of energy efficiency program implementation and evaluation. They can be considered the current primary resources for energy efficiency program impact evaluation and project measurement and verification (M&V), and thus the basis for the definitions, approaches, and issues adopted and explained in this guide. In addition, throughout the guide there are numerous documents referenced in endnotes—see individual guide sections for these references, as well as the references section that follows this appendix.

C.1 BACKGROUND

The information in this guide 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. Unfortunately, most of the early reference documents are not easily available to the general public (i.e., they are not posted on the Web). However, here are three reference documents in the public domain that provide a historical perspective and solid fundamentals:

- Hirst, E.; Reed, J. eds. (1991). *Handbook of Evaluation of Utility DSM Programs*. ORNL/CON-336. Prepared for Oak Ridge National Laboratory.
- Khawaja, M. S.; Mulholland, C.; Thayer, J.; Smith, K. (2008). *Guidebook for Energy Efficiency Program Evaluation Measurement & Verification.* 1016083. Palo Alto, CA: Electric Power Research Institute (EPRI).
- Violette, D. M. (1995). *Evaluation, Verification, and Performance Measurement of Energy Efficiency Programmes.* Prepared for International Energy Agency.

C.2 PRIMARY IMPACT EVALUATION RESOURCES

C.2.1 General Efficiency Evaluation Resource Websites

American Council for an Energy-Efficient Economy (ACEEE) publications: www.aceee.org/publications.

California Institute of Energy and Environment, Planning and Evaluation Library: http://uc-ciee.org/planning-evaluation/7/lbrsearch.

California Measurement Advisory Council (CALMAC): www.calmac.org.

Consortium for Energy Efficiency EM&V Resources: www.cee1.org/eval/eval-res.php3.

Efficiency Valuation Organization (EVO) International Performance Measurement and Verification Protocol (IPMVP): www.evo-world.org.

European Union Energy Efficiency Directive, measurement, monitoring, and evaluation: www.evaluate-energy-savings.eu/emeees/en/home/index.php.

Federal Energy Management Program M&V: http://ateam.lbl.gov/mv.

International Energy Program Evaluation Conference: www.iepec.org/IEPECHome.htm?programsabstracts.htm.

Northeast Energy Efficiency Partnerships EM&V Forum: http://neep.org/emv-forum.

Northwest Energy Efficiency Alliance evaluation: www.nwalliance.org/research/evaluationreports.aspx.

State and Local Energy Efficiency Action Network Resources: www.seeaction.energy.gov/resources.html.

U.S. Environmental Protection Agency EM&V Webinar Series: www.emvwebinar.org.

C.2.2 Select Impact Evaluation Resources

California Institute for Energy and Environment. (2009). *Energy Efficiency Evaluation Training Opportunities. http://uc-ciee.org/planning-evaluation/7/342/105/nested.*

California Public Utilities Commission. (2001). *California Standard Practice Manual: Economic Analysis of Demand-Side Programs and Projects.* www.energy.ca.gov/greenbuilding/documents/background/07-J_CPUC_STANDARD_PRACTICE_MANUAL.PDF. See also www.cpuc.ca.gov/PUC/energy/Energy+Efficiency/EM+and+V for the 2007SPM Clarification Memo.

California Public Utilities Commission. (2006). *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals.* Prepared by The TecMarket Works Team. www.calmac.org/events/EvaluatorsProtocols_Final_AdoptedviaRuling_06-19-2006.pdf.

California Public Utilities Commission. (2004). *California Evaluation Framework* (2004). Prepared by The TecMarket Works Team. www.calmac.org/publications/California_Evaluation_Framework_June_2004.pdf.

Messenger, M.; Bharvirkar, R.; Golemboski, B.; Goldman, C.A.; Schiller, S.R. (April 2010). *Review of Evaluation, Measurement and Verification Approaches Used to Estimate the Load Impacts and Effectiveness of Energy Efficiency Programs.* Lawrence Berkeley National Laboratory. Report LBNL-3277E. http://emp.lbl.gov/sites/all/files/lbnl-3277e.pdf.

State and Local Energy Efficiency Action Network. (April 2011). *National Energy Efficiency Evaluation, Measurement, and Verification (EM&V) Standard: Scoping Study of Issues and Implementation Requirements.* Prepared by Schiller, S.R.; Goldman, C.A.; Galawish, E.; LBNL Environmental Energy Technologies Division. www1.eere.energy.gov/seeaction/pdfs/emvstandard_scopingstudy.pdf.

State and Local Energy Efficiency Action Network. (June 2011). *Scoping Study to Evaluate Feasibility of National Databases for EM&V Documents and Measure Savings.* Prepared by Jayaweera, T.; Haeri, H.; Lee, A.; Bergen, S.; Kan, C.; Velonis, A.; Gurin, C.; Visser, M.; Grant, A.; Buckman, A.; The Cadmus Group Inc. www1.eere.energy.gov/seeaction/pdfs/emvscoping__databasefeasibility.pdf.

State and Local Energy Efficiency Action Network. (May 2012). *Evaluation, Measurement, and Verification (EM&V) of Residential Behavior-Based Energy Efficiency Programs: Issues and Recommendations.* Prepared by Todd, A.; Stuart, E.; Schiller, S.; Goldman, C.; Lawrence Berkeley National Laboratory. www1.eere.energy.gov/seeaction/pdfs/emv_behaviorbased_eeprograms.pdf.

U.S. Department of Energy. (2006). *EERE Guide for Managing General Program Evaluation Studies.* Prepared by H. Barnes, Lockheed Martin Aspen; Gretchen Jordan, Sandia National Laboratories. www1.eere.energy.gov/analysis/pdfs/evl_mg_app.pdf.

U.S. Department of Energy. (2007). *Impact Evaluation Framework for Technology Deployment Programs.* Prepared by J. Reed, Innovologie, LLC; G. Jordan, Sandia National Laboratories; E. Vine, Lawrence Berkeley National Laboratory. www1.eere.energy.gov/analysis/pdfs/impact_framework_tech_deploy_2007_overview.pdf.

U.S. Environmental Protection Agency. (2012). *The Roadmap for Incorporating Energy Efficiency/Renewable Energy Policies and Programs into State and Tribal Implementation Plans.* (See Appendix I of the roadmap). www.epa.gov/airquality/eere.

C.2.3 U.S. Environmental Protection Agency Non-Energy Benefits and Avoided Emissions Calculation References

Assessing the Multiple Benefits of Clean Energy

This document helps state energy, environmental, and economic policymakers identify and quantify the benefits of clean energy, including the energy, environmental—specifically greenhouse gas, air, and health—and economic benefits of clean energy. It provides an analytical framework that states can use to estimate those benefits during the development, implementation, and evaluation of clean energy policies and programs. http://epa.gov/statelocalclimate/resources/benefits.html.

Co-Benefits Risk Assessment (COBRA) Screening Model

EPA's Co-Benefits Risk Assessment (COBRA) screening model is a free tool that helps state and local governments estimate and map the air quality, human health, and related economic benefits (excluding energy cost savings) of clean energy policies or programs.

Roadmap for Incorporating Energy Efficiency/Renewable Energy Policies and Programs into State and Tribal Implementation Plans

The EPA Roadmap reduces the barriers for state, tribal, and local agencies to incorporate energy efficiency and renewable energy (EE/RE) policies and programs in state and tribal implementation plans by clarifying existing EPA guidance and providing new and detailed information. The goal of this document is to facilitate the use of energy efficiency and renewable energy emissions reduction strategies in air quality plans. www.epa.gov/airquality/eere/manual.html.

Projected Impacts of Existing State Energy Efficiency and Renewable Energy Policies

These EPA methods and projected energy impacts may be useful to states preparing State Implementation Plan (SIP) submittals to meet the National Ambient Air Quality Standards (NAAQS) for ozone and particulate matter. www.epa.gov/statelocalclimate/state/statepolicies.html.

Emissions & Generation Resource Integrated Database (eGRID)

eGRID is a comprehensive source of data on the environmental characteristics of almost all electric power generated in the United States. These include air emissions of nitrogen oxides, sulfur dioxide, carbon dioxide, and many other attributes. www.epa.gov/cleanenergy/energy-resources/egrid/index.html.

Power Plant Emissions Calculator (P-PEC)

P-PEC is a simplified tool that uses eGRID "non-baseload" emissions factors to help states quickly estimate the magnitude of emission reductions from EE/RE for each power plant within a region, and to understand potential emission reductions within a county or air quality nonattainment area. Contact Robyn DeYoung at EPA: deyoung.robyn@epa.gov

Hourly Marginal Emissions Tool

This tool is a statistical dispatch simulator that predicts the hourly changes in generation and air emissions at electric generating units (EGUs) resulting from EE/RE policies and programs. Contact Robyn DeYoung at EPA: deyoung.robyn@epa.gov

C.2.4 Technical Reference Manual (TRMs) Resources

TABLE C.1: United States Technical Reference Manuals

SCOPE	RESOURCE NAME	WEBSITE	FORMAT	INFORMATION INCLUDED	ADMINISTRATOR
National	ENERGY STAR®	www.energystar.gov/index.cfm?c=products.pr_find_es_products	Online Calculators	Ex ante savings based on algorithms	Agency
Regional – Northwest	Regional Technical Forum (RTF) Deemed Measures	www.nwcouncil.org/energy/rtf/measures/Default.asp	Online Database	Ex ante savings based on algorithms	Advisory Committee
Regional – Mid-Atlantic	Mid-Atlantic TRM	http://neep.org/uploads/EMV%20Forum/EMV%20Products/Mid%20Atlantic%20TRM_V1%202_FINAL.pdf	PDF	Algorithms and ex ante savings	Nonprofit Organization
Arkansas	Arkansas Deemed Savings Quick Start Programs	www.aepefficiency.com/oklahoma/ci/downloads/Deemed_Savings_Report.pdf	PDF	Algorithms	Public Utility
California	DEER Database for Energy-Efficient resources	www.deeresources.com	Software Program	Ex ante savings	State Commission
Connecticut	Connecticut Light & Power and United Illuminating Company Program Savings Documentation	www.ctenergyinfo.com/2012 CT Program Savings Documentation FINAL.pdf	PDF	Algorithms and ex ante savings	Public Utility
Hawaii	Hawaii Energy Efficiency Program TRM	www.hawaiienergy.com/media/assets/AnnualReportAttachments-ALL.pdf	PDF	Algorithms and ex ante savings	State Commission
Maine	Efficiency Maine TRM – Commercial	www.efficiencymaine.com/docs/board_meeting_documents/Maine-Commercial-TRM-8-31-2010-Final.pdf	PDF	Algorithms and ex ante savings	Trust
	Efficiency Maine TRM – Residential	www.efficiencymaine.com/docs/board_meeting_documents/Maine-Residential-TRM-02-04-09.pdf	PDF		
Massachusetts	Massachusetts Statewide TRM for Estimating Savings from Energy Efficiency Measures	www.ma-eeac.org/docs/2012 MTM Files/Oct 28, 2011 DPU Filings/WMECo/13 - WMECO Exhibit H MA TRM_2012 PLAN_FINAL.pdf	PDF	Algorithms and ex ante savings	Agency

SCOPE	RESOURCE NAME	WEBSITE	FORMAT	INFORMATION INCLUDED	ADMINISTRATOR
Michigan	Michigan Energy Measures Database	www.michigan.gov/mpsc/0,1607,7-159-52495_55129---,00.html	Excel Database	Ex ante savings	State Commission
New Jersey	New Jersey Clean Energy Program Protocols to Measure Resource Savings	www.njcleanenergy.com/files/file/Library/NJ Protocols Revisions 7-21-11_Clean.pdf	PDF	Algorithms and ex ante savings	Agency
New York	New York Standard Approach for Estimating Energy Savings from Energy Efficiency Programs	http://efile.mpsc.state.mi.us/efile/docs/16671/0026.pdf	PDF	Algorithms and ex ante savings	Agency
Ohio	Ohio TRM	www.ohiotrm.org	Online Database	Algorithms and ex ante savings	State Commission
Pennsylvania	Pennsylvania TRM	www.puc.state.pa.us/electric/Act129/TRM.aspx	DOC	Algorithms and ex ante savings	State Commission
Texas	Deemed Savings, Installation, and Efficiency Standards	www.entergy-texas.com/content/Energy_Efficiency/documents/Deemed_Savings_Measures_List.pdf	PDF	Algorithms and ex ante savings	State Commission
Vermont	Efficiency Vermont Technical Reference User Manual	www.veic.org/Libraries/Resumes/TechManualEVT.sflb.ashx	PDF	Algorithms and ex ante savings	Nonprofit Organization
Wisconsin	Focus on Energy Evaluation Business Programs: Deemed Savings Manual V1.0	www.focusonenergy.com/files/Document_Management_System/Evaluation/bpdeemedsavings-manuav10_evaluationreport.pdf	PDF	Algorithms and ex ante savings	State Commission

Source: State and Local Energy Efficiency Action Network. (June 2011). Scoping Study to Evaluate Feasibility of National Databases for EM&V Documents and Measure Savings. Prepared by Jayaweera, T.; Haeri, H.; Lee, A.; Bergen, S.; Kan, C.; Velonis, A.; Gurin, C.; Visser, M.; Grant, A.; Buckman, A.; The Cadmus Group Inc. www1.eere.energy.gov/seeaction/pdfs/emvscoping__databasefeasibility.pdf. Tables are current as of Summer 2012.

C.3 MEASUREMENT AND VERIFICATION RESOURCES

Several M&V resource documents used in the development of this guide are available via the internet and are presented in this section; they can be considered the current primary resources for energy efficiency project M&V.

- **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 the nonprofit Efficiency Valuation Organization, which continually maintains and updates the IPMVP.

  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 M&V for specific projects; 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), presented in Chapter 4, differentiate the most common approaches for M&V. Reference: Efficiency Valuation Organization (2010). International Performance Measurement and Verification Protocol. www.evo-world.org.

- **FEMP M&V Guidelines.** The purpose of this document is to provide guidelines and methods for documenting 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 operations and maintenance, renewable energy, and cogeneration projects. Along with the FEMP M&V Guidelines are several useful companion documents. http://mnv.lbl.gov/keyMnVDocs.

- **ASHRAE Guideline 14 Measurement of Energy and Demand Savings.** ASHRAE is the professional engineering society that has been the most involved in writing guidelines and standards associated with energy efficiency. Compared with 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). A new version is expected to be available in 2013. www.ashrae.org.

- **ASHRAE Performance Measurement Protocols (PMP) for Commercial Buildings.** These provide a standardized, consistent set of protocols for facilitating the comparison of measured energy, water, and indoor quality performance of commercial buildings. www.ashrae.org.

- **International Standards Organization (ISO).** The International Standards Organization entered the world of energy management with the release of ISO 50001 - Energy Management Systems. This protocol includes the following:
  - ISO 50001:2011 specifies requirements for establishing, implementing, maintaining, and improving an energy management system, the purpose of which is to enable an organization to follow a systematic approach in achieving continual improvement of energy performance, including energy efficiency, energy use, and consumption.
  - ISO 50001:2011 specifies requirements applicable to energy use and consumption, including measurement, documentation and reporting, design, and procurement practices for equipment, systems, processes, and personnel that contribute to energy performance.
  - ISO 50001:2011 applies to all variables affecting energy performance that can be monitored and influenced by the organization. ISO 50001:2011 does not prescribe specific performance criteria with respect to energy.
  - Subsequent to the release of 50001, ISO created a technical committee (TC242) to support the deployment of ISO 50001. In January 2011, the ISO Technical Management Board announced the creation of another committee to create standards on measuring and verifying savings—TC 257 Energy Savings. Any standards originating from TC 242, TC 257, or the joint working group are expected to provide broad guidance on M&V principles in 2012 or 2013.

References

"Acid Rain Program." (2012). U.S. Environmental Protection Agency. www.epa.gov/airmarkets/progsregs/arp.

"Air Markets Program Data." (2012). U.S. Environmental Protection Agency. http://ampd.epa.gov/ampd/.

American Council for an Energy-Efficient Economy (ACEEE). (October 2011). *State Energy Efficiency Resource Standard (EERS) Activity.* http://aceee.org/files/pdf/policy-brief/. State EERS Summary October 2011.pdf.

American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). (2002). Guideline 14: *Measurement of Demand and Energy Savings.* www.ashrae.org.

American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). (2010). *Performance Measurement Protocols for Commercial Buildings.* www.ashrae.org/standards-research--technology/special--project-activities.

Barbose, G.; Billingsley; M.; Goldman, C.; Hoffman, I.; Schlegel, J. (August 2012). "On a Rising Tide: The Future of U.S. Utility Customer-Funded Energy Efficiency Programs." *2012 ACEEE Summer Study Proceedings;* August 12-17, 2012, Pacific Grove, California. Washington, D.C.: American Council for an Energy-Efficient Economy (ACEEE). LBNL-5755E. http://www.aceee.org/files/proceedings/2012/data/papers/0193-000173.pdf.

"Building Energy Codes Program." (2012). U.S. Department of Energy. www.energycodes.gov.

"California Greenhouse Gas (GHG) Cap-and-Trade Program." (2012). California Air Resources Board (CARB). www.arb.ca.gov/cc/capandtrade/capandtrade.htm.

California Public Utilities Commission (CPUC). (2001). *California Standard Practice Manual: Economic Analysis of Demand-Side Programs and Projects.* www.energy.ca.gov/greenbuilding/documents/background/07-J_CPUC_STANDARD_PRACTICE_MANUAL.PDF. See also http://www.cpuc.ca.gov/PUC/energy/Energy+Efficiency/EM+and+V/.

California Public Utilities Commission (CPUC). (April 2006). *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals.* www.calmac.org/events/EvaluatorsProtocols_Final_AdoptedviaRuling_06-19-2006.pdf.

California Public Utilities Commission (CPUC). (2007). *Guidelines for Estimating Net-To-Gross Ratios Using the Self-Report Approaches.* CPUC Energy Division Master Evaluation Contractor Team. ftp://ftp.cpuc.ca.gov/puc/energy/electric/energy+efficiency/ee+workshops/selfreportguidelinesdetailed_v20.pdf.

California Public Utilities Commission (CPUC). (April 9, 2010). *Codes & Standards (C&S) Programs Impact Evaluation: California Investor Owned Utilities' Codes and Standards Program Evaluation for Program Years 2006–2008.* CALMAC Study ID: CPU0030.06 Final Evaluation Report. Prepared by KEMA, Inc.; The Cadmus Group, Inc.; Itron, Inc.; Nexus Market Research, Inc.; and ENRG, Inc. for the CPUC Energy Division. www.calmac.org/publications/Codes_Standards_Vol_III_FinalEvaluationReportUpdated_04122010.pdf.

"Clean Air Interstate Rule." (2005). U.S. Environmental Protection Agency. www.epa.gov/cair.

"Compilation of Air Pollutant Emission Factors." (2011). U.S. Environmental Protection Agency. EPA AP-42. www.epa.gov/ttn/chief/ap42.

Consensus Report on the EmPower Maryland EM&V Process. (June 24, 2009). www.raponline.org/document/download/id/551.

Efficiency Valuation Organization (EVO). International Performance Measurement and Verification Protocol (IPMVP). (multiple dates). www.evo-world.org.

"eGRID." (2012). U.S. Environmental Protection Agency. www.epa.gov/cleanenergy/energy-resources/egrid/index.html.

Energy & Resources Solutions, Inc. (ERS); Left Fork Energy; West Hill Energy & Computing; Megdal & Associates, LLC.; Opinion Dynamics Corporation. (June 2010). *Report on Impact Assessments of Two NYSERDA Natural Gas Efficiency Programs: Con Edison Natural Gas Efficiency Program & National Grid Low Income Natural Gas Efficiency Program.* Prepared for the New York State Energy Research and Development Authority (NYSERDA). http://documents.dps.ny.gov/public/Common/ViewDoc.aspx?DocRefId={CFCD7DE1-8B68-46CC-B64B-83ACD072A4FD}.

Eto, J.; Prahl, R.; Schegel, J. (1996). *A Scoping Study on Energy-Efficiency Market Transformation by California Utility DSM Programs*. Lawrence Berkeley National Laboratory. http://eetd.lbl.gov/ea/ems/reports/39058.pdf.

Federal Energy Management Program (FEMP). (April 2008). *M&V Guidelines: Measurement and Verification for Federal Energy Projects.* www1.eere.energy.gov/femp/pdfs/mv_guidelines.pdf.

Federal Energy Regulatory Commission (FERC). (June 2010). *National Action Plan on Demand Response.* Docket No. AD09-10. Prepared by FERC staff. www.ferc.gov/legal/staff-reports/06-17-10-demand-response.pdf.

"Global Warming Solutions Act: Assembly Bill 32." (2006). California Air Resources Board (CARB). www.arb.ca.gov/cc/ab32/ab32.htm.

Gottstein, M.; Schwartz, L. (2010). T*he Role of Forward Capacity Markets in Increasing Demand-Side and Other Low-Carbon Resources Experience and Prospects.* Regulatory Assistance Project (RAP). www.raponline.org/docs/RAP_Gottstein_Schwartz_RoleofFCM_ExperienceandProspects2_2010_05_04.pdf.

Haeri, H.; Khawaja, M.S. (March 2012). "The Trouble With Freeriders." *Public Utilities Fortnightly.* www.fortnightly.com/fortnightly/2012/03/trouble-freeriders.

Hibbard, P.; Tierney, S.; Okie, A.; Darling, P. (November 2011). *The Economic Impacts of the Regional Greenhouse Gas Initiative on Ten Northeast and Mid-Atlantic States, Review of the Use of RGGI Auction Proceeds from the First Three-Year Compliance Period.* The Analysis Group. www.analysisgroup.com/uploadedfiles/publishing/articles/economic_impact_rggi_report.pdf.

Hirst, E.; White, D. (1985). *Indoor Temperature Changes After Retrofit: Inferences Based on Electricity Billing Data for Nonparticipants and Participants in the BPA Residential Weatherization Program*. Oak Ridge National Laboratory. www.ornl.gov/info/reports/1985/3445600444954.pdf.

Intergovernmental Panel on Climate Change (IPCC). (2006). *2006 Guidelines for National Greenhouse Gas Inventories.* www.ipcc-nggip.iges.or.jp/public/2006gl/index.html.

Intergovernmental Panel on Climate Change (IPCC). (2007). *Climate Change 2007: Synthesis Report*. www.ipcc.ch/pdf/assessment-report/ar4/syr/ar4_syr.pdf.

Intergovernmental Panel on Climate Change (IPCC). (2007). Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, 2007. www.ipcc.ch/publications_and_data/ar4/wg3/en/contents.html.

Intergovernmental Panel on Climate Change (IPCC). (2007). *Pathways to a Low-Carbon Economy: Version 2 of the Global Greenhouse Gas Abatement Cost Curve.* Prepared by McKinsey & Company. www.epa.gov/statelocalclimate/documents/pdf/mckinsey_summary_11-19-09.pdf.

Kushler, M.; Nowak, S.; Witte, P. (February 2012). *A National Survey of State Policies and Practices for the Evaluation of Ratepayer-Funded Energy Efficiency Programs.* American Council for an Energy-Efficient Economy (ACEEE). Report Number U122. www.aceee.org/research-report/u122.

Lazar, J.; Baldwin, X. (August 2011). *Valuing the Contribution of Energy Efficiency to Avoided Marginal Line Losses and Reserve Requirements.* Regulatory Assistance Project (RAP). www.raponline.org/document/download/id/4537.

Mahone, D.; Hall, N. (August 2010). "Pursuit of Aggressive Energy Savings Targets—Aligning Program Design and Evaluation." *2010 ACEEE Summer Study Proceedings;* August 15-20, 2010, Pacific Grove, California. Washington, D.C.: American Council for an Energy-Efficient Economy. http://neep.org/uploads/EMV%20Forum/Steering%20Committee%20Notes/DMahone_ACEEE_Paper_for_SC_Notes.pdf.

Mahone, D. (August 2011). "Free-Ridership as a Way to Kill Programs—How Evaluation Policies Can Frustrate Efficiency Goals." Boston: International Energy Program Evaluation Conference. www.iepec.org/2011PapersTOC/papers/054.pdf.

Messenger, M.; Bharvirkar, R.; Golemboski, B.; Goldman, C.A.; Schiller, S.R. (April 2010). *Review of Evaluation, Measurement and Verification Approaches Used to Estimate the Load Impacts and Effectiveness of Energy Efficiency Programs.* Lawrence Berkeley National Laboratory. Report LBNL-3277E. http://emp.lbl.gov/sites/all/files/lbnl-3277e.pdf.

Mid-Atlantic Regional Air Management Association (MARAMA). www.marama.org.

National Action Plan for Energy Efficiency. (2007). *Guide to Resource Planning with Energy Efficiency.* Prepared by Snuller Price et al., Energy and Environmental Economics, Inc. (E3). www.epa.gov/cleanenergy/documents/suca/resource_planning.pdf.

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.

National Action Plan for Energy Efficiency. (2008). *Understanding Cost-Effectiveness of Energy Efficiency Programs: Best Practices, Technical Methods, and Emerging Issues for Policy-Makers.* Prepared by Energy and Environmental Economics, Inc. (E3); Regulatory Assistance Project (RAP). www.epa.gov/cleanenergy/documents/suca/cost-effectiveness.pdf.

NMR Group, Inc.; Research Into Action, Inc. (November 2010). "Net Savings Scoping Paper." Submitted to Northeast Energy Efficiency Partnerships: Evaluation, Measurement, and Verification Forum. http://neep.org/uploads/EMV%20Forum/EMV%20Products/FINAL%20Net%20Savings%20Scoping%20Paper%2011-13-10.pdf.

Nowak, S.; Kushler, M.; Sciortino, M.; York, D.; Witte, P. (June 2011). *Energy Efficiency Resource Standards: State and Utility Strategies for Higher Energy Savings.* American Council for an Energy-Efficient Economy (ACEEE). Report Number U113. www.aceee.org/research-report/u113.

Peters, J.; McRae, M. (August. 2008). "Free-Ridership Measurement Is Out of Sync With Program Logic...or, We've Got the Structure Built, but What's Its Foundation?" *2008 ACEEE Summer Study Proceedings.* August 17-22, 2008, Pacific Grove, California. Washington, DC: American Council for an Energy-Efficient Economy (ACEEE). www.aceee.org/files/proceedings/2008/data/papers/5_491.pdf.

"Promoting Air Quality and Climate Protection with Clean Energy and Energy Efficiency in the Metropolitan Washington Area." Metropolitan Washington Council of Governments (MWCOG). www.mwcog.org/environment/air/EERE/default.asp.

Rathbun, P.; Sabo, C.; Zent, B. (2003). *Standardized Methods for Free-Ridership and Spillover Evaluation—Task 5 Final Report* (Revised). Prepared for National Grid, NSTAR Electric, Northeast Utilities, Unitil, and Cape Light Compact. www.cee1.org/eval/db_pdf/297.pdf.

"Regional Greenhouse Gas Initiative." (2012). http://rggi.org/.

Rosenberg, M.; Hoefgen, L. (March 2009). *Market Effects and Market Transformation: Their Role in Energy Efficiency Program Design and Evaluation.* Prepared for California Institute for Energy and Environment (CIEE) by KEMA, Inc.; Nexus Market Research. http://uc-ciee.org/downloads/mrkt_effts_wp.pdf.

Schiller, S.R. (August 2011). "Documenting Greenhouse Gas Emission Reductions from End-Use Efficiency Activities in the Kyoto Protocol's Clean Development Mechanism Offset Program." *2011 International Energy Program Evaluation Conference Proceedings,* August 16-18, 2011, Boston, Massachusetts. Madison, Wisconsin: International Energy Program Evaluation Conference (IEPEC). www.iepec.org/2011PapersTOC/papers/115.pdf#page=1.

Schiller, S. R.; Goldman, C.A. (August 2011). "Developing State and National Evaluation Infrastructures—Guidance for the Challenges and Opportunities of EM&V." *2011 International Energy Program Evaluation Conference Proceedings,* August 16-18, 2011, Boston, Massachusetts. Madison, Wisconsin: International Energy Program Evaluation Conference (IEPEC). www.iepec.org/2011PapersTOC/papers/073.pdf#page=1.

Schiller, S.R.; Prindle, B.; Cowart, R.; Rosenfeld, A.H. (2008). "Energy Efficiency and Climate Change Mitigation Policy." *2008 ACEEE Summer Study Proceedings.* August 17-22, 2008, Pacific Grove, California. Washington, DC: American Council for an Energy-Efficient Economy (ACEEE). www.aceee.org/files/proceedings/2008/data/papers/8_306.pdf.

Skumatz, L.A. (1997). "Recognizing All Program Benefits: Estimating the Non-Energy Benefits of PG&E's Venture Partners Pilot Program (VPP)." *Proceedings of the 1997 Energy Evaluation Conference.* August 27-29, 1997, Chicago, Illinois. Madison, Wisconsin: International Energy Program Evaluation Conference (IEPEC). www.iepec.org/1997PapersTOC/papers/033.pdf.

Skumatz, L.A.; Khawaya, M.; Colby, J. (2009). *Lessons Learned and Next Steps in Energy Efficiency Measurement and Attribution: Energy Savings, Net to Gross, Non-Energy Benefits, and Persistence of Energy Efficiency Behavior.* Prepared for the California Institute for Energy and Environment (CIEE) Behavior and Energy Program. http://uc-ciee.org/downloads/EEM_A.pdf.

Skumatz, L.A.; Khawaja, M.S.; Krop, R. (May 2010). *Non-Energy Benefits: Status, Findings, Next Steps, and Implications for Low Income Program Analyses in California.* Revised Report. Prepared for Brenda Gettig, Sempra Utilities. www.liob.org/docs/LIEE Non-Energy Benefits Revised Report.pdf.

State and Local Energy Efficiency Action Network. (April 2011). *National Energy Efficiency Evaluation, Measurement, and Verification (EM&V) Standard: Scoping Study of Issues and Implementation Requirements.* Prepared by Schiller, S.R.; Goldman, C.A.; Galawish, E.; LBNL Environmental Energy Technologies Division. www1.eere.energy.gov/seeaction/pdfs/emvstandard_scopingstudy.pdf.

State and Local Energy Efficiency Action Network. (June 2011). *Scoping Study to Evaluate Feasibility of National Databases for EM&V Documents and Measure Savings.* Prepared by Jayaweera, T.; Haeri, H.; Lee, A.; Bergen, S.; Kan, C.; Velonis, A.; Gurin, C.; Visser, M.; Grant, A.; Buckman, A.; The Cadmus Group Inc. www1.eere.energy.gov/seeaction/pdfs/emvscoping__databasefeasibility.pdf.

State and Local Energy Efficiency Action Network. (May 2012). *Evaluation, Measurement, and Verification (EM&V) of Residential Behavior-Based Energy Efficiency Programs: Issues and Recommendations.* Prepared by Todd, A.; Stuart, E.; Schiller, S.; Goldman, C.; Lawrence Berkeley National Laboratory. www1.eere.energy.gov/seeaction/pdfs/emv_behaviorbased_eeprograms.pdf.

Stewart, J.; Haeri, M.H. (July 2011). *Critical Review and Recommendations on Top-Down Evaluation.* White paper. The Cadmus Group, Inc. Prepared for the California Public Utilities Commission (CPUC).

"Uniform Methods Project." (2012). U.S. Department of Energy. www1.eere.energy.gov/deployment/ump.html.

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.* www.epa.gov/ttn/oarpg/t1/memoranda/ereseerem_gd.pdf.

U.S. Environmental Protection Agency. (September 2005). (Draft). *State Set-Aside Programs for Energy Efficiency and Renewable Energy Projects Under the NOx Budget Trading Program: A Review of Programs in Indiana, Maryland, Massachusetts, Missouri, New Jersey, New York, and Ohio.* http://epa.gov/statelocalclimate/documents/pdf/eere_rpt.pdf.

U.S. Environmental Protection Agency. (February 2010). *Assessing the Multiple Benefits of Clean Energy: A Resource for States.* www.epa.gov/statelocalclimate/resources/benefits.html.

U.S. Environmental Protection Agency. (August 2011). *Federal Implementation Plans: Interstate Transport of Fine Particulate Matter and Ozone and Correction of SIP Approvals.* www.gao.gov/products/C00127. 40 CFR Parts 51, 52, 72, 78, and 97 [EPA–HQ–OAR–2009–0491; FRL–9436–8] RIN 2060–AP50. ACTION: Final rule.

Vine, E.; Hall, N.; Keating, K.M.; Kushler, M.; Prahl, R. (August 2011). *"Emerging Evaluation Issues Revisited." 2011 International Energy Program Evaluation Conference Proceedings,* August 16-18, 2011, Boston, Massachusetts. Madison, Wisconsin: International Energy Program Evaluation Conference (IEPEC). www.iepec.org/2011PapersTOC/papers/096.pdf#page=1.

Wallace, P.; Forster, H.J. (2012). *State of the Efficiency Program Industry Budgets, Expenditures, and Impacts 2011.* Consortium for Energy Efficiency. www.cee1.org/files/2011 CEE Annual Industry Report.pdf.

Woolf, T.; Malone, E.; Takahashi, K.; Steinhurst, W. (July 23, 2012). *Best Practices in Energy Efficiency Program Screening: How to Ensure that the Value of Energy Efficiency is Properly Accounted For. Synapse Energy Economics, Inc.* Prepared for National Home Performance Council. www.synapse-energy.com/Downloads/SynapseReport.2012-07.NHPC.EE-Program-Screening.12-040.pdf.

World Resources Institute (WRI); World Business Council for Sustainable Development (WBCSD). (December 2005). *GHG Protocol for Project Accounting* (Project Protocol). www.ghgprotocol.org/files/ghgp/ghg_project_protocol.pdf.

World Resources Institute (WRI); World Business Council for Sustainable Development (WBCSD). (August 2007). *Protocol Guidelines for Quantifying GHG Reductions from Grid-Connected Electricity Projects.*www.wri.org/publication/guidelines-quantifying-ghg-reductions-grid-connected-electricity-projects.

Zabin, C. (2011). *California Workforce Education and Training Needs Assessment.* Center for Labor Research & Education, Donald Vial Center on Employment in the Green Economy. www.irle.berkeley.edu/vial.

SEE Action

STATE & LOCAL ENERGY EFFICIENCY ACTION NETWORK