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

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

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

2.5 Notes

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

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

Ameren Ex 7.2

3: Impact Evaluation Basics

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

3.1 Impact Evaluation Process

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

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

Basic Impact Evaluation Concepts

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

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

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

3.2 How Energy and Demand Savings Are Determined

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

1. Gross program energy and demand savings are determined.

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

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

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

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

Quality Assurance Guidelines

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

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

Avoided emissions are calculated by applying emission factors (for example, pounds of carbon dioxide per kWh of savings) to the net energy savings value. What constitutes net energy savings for an avoided emissions program—along with the sources of emission factors—is discussed in Section 3.9 and Chapter 6.

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

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

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

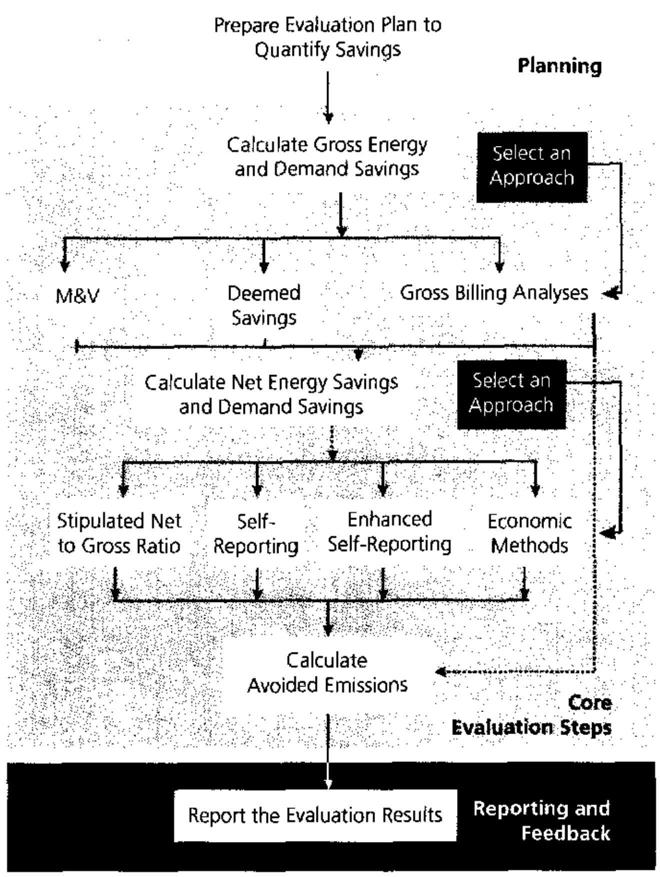
1. Defining evaluation goals and scale, including deciding which program benefits to evaluate.
2. Setting the time frame for the evaluation and reporting expectations.
3. Setting a spatial boundary for evaluation (i.e., what energy uses, emission sources, etc., will be included in the analyses).
4. Defining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

3.2.1 Approaches for Calculating Gross Energy and Demand Savings

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

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

Figure 3-1. The Impact Evaluation Process



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

Estimation of Gross Energy Savings

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

Estimation of Net Energy Savings

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

M&V Versus Deemed Savings

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

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

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

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

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

3.2.2 Approaches for Calculating Net Energy and Demand Savings

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

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

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

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

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

It is not unusual for combinations of these approaches to be used. For example, rigorous econometric methods may be used every three years and self-reported or deemed NTGRs are used for the other program years. If a previous econometric study is considered more reliable, its result may be used as the deemed value or the self-reported calculations may be calibrated to come closer to the previous result.

National Grid Net Savings Example

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

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

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

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

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

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

3.3 Calculating Demand Savings

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

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

Regional Coincident Peak Demand

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

Examples of demand savings definitions are:

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

what reduction occurs when there is a call for demand reductions. The evaluation can be of the: (a) level of demand reduction that has been pledged or enabled through testing and inspection or (b) level of demand reduction that has been achieved using a variety of methods, some of which are included in this Guide and some of which are specific to demand response.

The calculation for demand savings is straightforward:

demand savings = energy savings / time period of energy savings (eq 3.1)

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

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

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

Secondary methods rely upon collected energy consumption data that are only available as averaged values over longer periods, such as monthly or even annually. When longer periods are used, demand impacts can also be estimated from energy impacts by applying a series of standard *load shapes* to allocate energy consumption into costing period bins. These

Demand Response Evaluation

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

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

load shapes (for whole facilities or by end-use) may be available from other studies for related programs in similar markets. One source for the load shape data is the energy savings load shapes, by measure, that are included in the California Database for Energy Efficiency Resources (DEER).[4]

3.4 Co-Benefits

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

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

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

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

However, cost-effectiveness analysis requires that *at least* the most important types of benefits and costs be

Evaluating Participant Non-Energy Benefits

NEBs can be evaluated through a range of survey approaches:

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

All of these approaches have benefits and drawbacks. The industry standard, to date, has been CV and DQ approaches. However, in recent years, NYSERDA has pioneered the joint use of DQ and CA survey methods on its New York Energy $mart Program. Thus far, the DQ and CA approaches have resulted in individual NEB values within the same general range (note that NYSERDA uses the term "non-energy indicators"). However, values derived by CA fall toward the lower end of the range. This could be due to many factors, not the least of which is the more limited set of non-energy co-benefits that can reasonably be covered in CA surveys. Reference: NYSERDA, 2006.

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

July 1, 2001–December 31, 2006

Program Area	Value of Non-Energy Benefits FY07 as of December 31, 2006	Value of Non-Energy Benefits Program to Date as of December 31, 2006
Business Programs	$1.6 million	$15.2 million
Example Benefits from Business Programs: • Maintenance employee morale • Equipment life • Productivity • Waste generation	• Defects and errors • Sales • Non-energy costs • Personnel needs • Injuries and illnesses	
Residential Programs	$2.3 million	$24.1 million
Example Benefits from Residential Programs: • Increased safety resulting from a reduction of gases such as carbon monoxide due to the installation of a new high-efficiency furnace • Fewer illnesses resulting from elimination of mold problems due to proper air sealing, insulating and ventilation of a home • Reduced repair and maintenance expense due to having newer, higher quality equipment • Increased property values resulting from installation of new equipment • Reduced water and sewer bill from installation of a horizontal-axis washing machine, which uses much less water than a conventional washing machine		
Renewable Energy Programs	$0	$563,000
Example Benefits from Renewable Energy Programs: • Greater diversity of primary in-state energy supplies • Use of wastes as a fuel instead of disposal • Increased ability to handle energy emergencies or generation shortfalls • Increased sales of renewable energy byproducts		

Method of applying value is under review.

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

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

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

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

3.5 Persistence

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

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

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

The basic approaches for assessing persistence are:

- Use of historical and documented persistence data, such as manufacturer's studies or studies done by industry organizations such as ASHRAE.
- Laboratory and field testing of the performance of energy-efficient and baseline equipment.

- Field inspections, over multiple years, of efficiency activities that constitute a program.
- Non-site methods such as telephone surveys and interviews, analysis of consumption data, or use of other data (e.g., data from a facility's energy management system).

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

3.6 Uncertainty

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

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

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

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

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

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

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

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

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

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

3.7 Appropriate Applications of Impact Evaluations

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

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

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

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

3.8 Evaluation Characteristics and Ethics

Ideally, any evaluation process will be defined by the following principles.

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

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

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

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

3.9 Calculating Avoided Emissions

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

3.9.1 Energy Efficiency and Avoided Emissions

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

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

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

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

   $$\text{avoided emissions}_t = (\text{net energy savings})_t \times (\text{emission factor})_t \qquad (eq\ 3.2)$$

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

   The basic equation for this approach is:

   $$\text{avoided emissions} = (\text{base case emissions}) - (\text{reporting period emissions}) \qquad (eq\ 3.3)$$

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

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

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

3.9.2 Additionality

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

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

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

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

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

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

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

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

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

3.9.4 Special Issues for Capped Pollutants Under Cap and Trade Programs

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

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

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

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

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

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

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

3.9.5 Avoided Emissions Calculations for Different Objectives

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

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

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

The following documents provide some guidance on these issues, with respect to greenhouse gas programs. They were all prepared by the World Business Council for Sustainable Development (WBCSD) and/or the World Resources Institute (WRI) and are available at <http://www.wri.org/climate/>.

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

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

3.10 Notes

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

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

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

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

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

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

Ameren Ex 7.2

4: Calculating Gross Energy and Demand Savings

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

4.1 Basics of Calculating Gross Savings

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

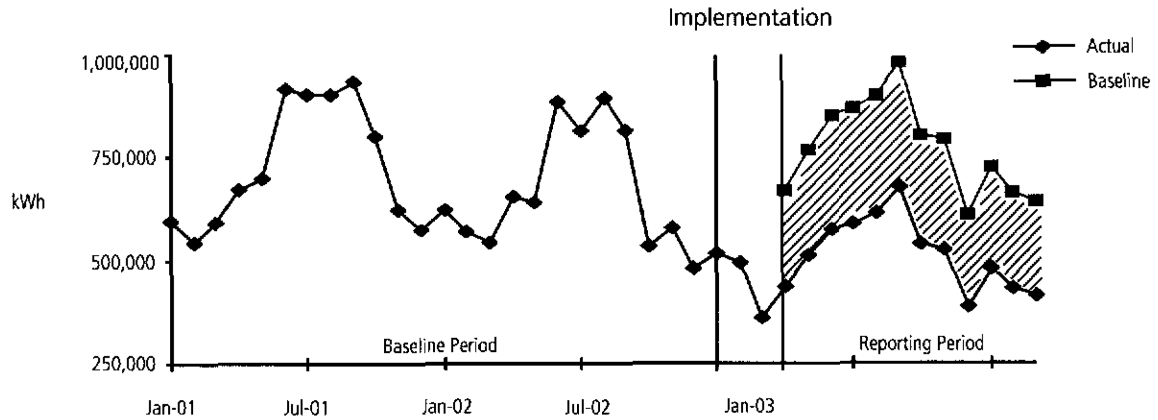
energy savings = (baseline energy use) – (reporting period energy use) ± (adjustments) *(eq 4.1)*

Weather Adjustments

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

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

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



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

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

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

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

4.2 Measurement and Verification Approach

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

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

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

Field Inspections of Energy Efficiency Measures

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

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

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

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

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

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

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

In terms of type of measurement:

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

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

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

Source: EVO, 2007.

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

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

Source: EVO, 2007.

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

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

Ameren Ex 7.2

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

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

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

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

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

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

Interactive Factors

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

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

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

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

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

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

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

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

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

Retrofit Isolation and Measurement Boundaries Example

A factory's boiler, used for process steam production, is replaced with a more efficient boiler of about the same capacity. The *measurement boundary* is defined to just include the boiler, whether the baseline boiler (before it is replaced) or the more efficient boiler (once it is installed). With this boundary, the analyses of baseline and efficient boilers are not affected by variations in the factory's process steam load, although the actual savings depend on the steam consumption of the factory. Meters for fuel consumption and boiler steam output are all that are needed to assess the efficiencies of the baseline and efficient boilers over their full range of operations. Under Option A, *savings* are reported for the boiler retrofit by applying the measured annual average efficiency improvement to an estimated annual boiler load and the boiler efficiency test is repeated annually during the reporting period. Under Option B, the annual boiler load may be determined by measuring the boiler load over several weeks (to prepare typical hourly and daily load profiles) and then using this information to make a more accurate savings estimate based on matching typical hourly load profiles to partial and full steam load boiler efficiency profiles, rather than just using an annual average efficiency value and an average annual steam consumption value.

4.2.3 M&V Option C—Whole-Facility Analyses

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

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

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

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

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

EPA's Portfolio Manager

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

4.2.4 M&V Option D—Calibrated Simulation

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

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

Building Energy Simulation Programs

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

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

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

4.3 Deemed Savings Approach

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

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

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

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

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

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

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

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

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

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

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

4.4 Large-Scale Data Analysis Approach

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

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

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

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

  $$savings = Q_{pre\text{-}installation} - Q_{post\text{-}installation} \quad (eq\ 4.1)$$

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

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

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

  $$savings = Q_{non\text{-}participants} - Q_{participants} \quad (eq\ 4.2)$$

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

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

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

  $$savings = (Q_{pre\text{-}installation} - Q_{post\text{-}installation})_{participants} - (Q_{pre\text{-}installation} - Q_{post\text{-}installation})_{non\text{-}participants} \quad (eq\ 4.3)$$

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

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

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

The primary consideration for any evaluation is that the analysis must be designed to obtain reliable energy savings. Uncertainty of savings estimates can decrease as the evaluators attempt to incorporate the major independent variables that may have affected the observed change in consumption. This can be accomplished in several ways. One common method is to include participant and non-participant analyses (the second and third bullets above). If one of these approaches is selected, particular care and justification must be made for the non-participant group selected and its appropriateness for the program and participant population being analyzed. Secondly, evaluation design and analysis needs to consider whether the analysis is providing gross impact, net impact, or something in between that must then be adjusted or analyzed.

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

In regression analysis, the following factors need to be addressed:

- What independent variables are relevant to calculating energy savings? Often this is decided by common sense, experience, or budget considerations (with respect to how many variables can be measured and tracked) but it also can be determined through field experiments and statistical tests. For weather data, the most common independent variable, there is a wide range of public and private data sources.
- Will a comparison group be used in the analysis? While often a more accurate approach, the use of comparison groups assumes that a comparable group of participants and non-participants can be found and analyzed. This, of course, adds to evaluation costs.
- How will the analysis be tested for statistical errors, and what level of uncertainty is acceptable? The first concern requires qualified analysts and a quality control system. The second requires specification of statistical parameters that define the uncertainty of the calculated savings. The field of statistical analysis can be quite complex and untrained analysts often misinterpret analyses and miss key considerations or errors in statistical analyses.
- Are gross or net savings values desired? The latter two methods described above, which include comparison groups, can actually produce net savings values.

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

- **Normalized annual consumption (NAC) analysis.** This is a regression-based method that analyzes monthly energy consumption data. The NAC analysis can be conducted using statistical software, such as the Princeton Scorekeeping Method (PRISM), and other statistically based approaches using SAS or SPSS.[6] The NAC method, often using PRISM, has been most often used to estimate energy impacts produced by whole-house retrofit programs.
- **Conditional savings analysis (CSA).** CSA is a type of analysis in which change in consumption is modeled using regression analysis against the presence or absence of energy efficiency measures. These are usually entered in the form of binary variables (1 if measures are installed and 0 if not).
- **Statistically adjusted engineering (SAE) models.** A category of statistical analysis models that incorporate the engineering estimate of savings as a dependent variable. For example, a SAE model can use change in energy as the dependant variable

in a regression model against estimated savings for installed efficiency measures. Often these estimates are provided in the design phase or through secondary sources (e.g., DEER). When the measures are installed, the estimated savings is entered as the explanatory variable value. When the measures are not installed, 0 is entered as the explanatory variable value in the regression model.

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

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

4.5 Selecting a Gross Savings Evaluation Approach

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

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

4.5.1 Large-Scale Data Analysis Approach

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

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

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

4.5.2 Deemed Savings Approach

Deemed savings approaches are most commonly used for programs that involve simple new construction or retrofit energy efficiency measures with well-defined