APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R-Doc. 133

ARKANSAS PUBLIC SERVICE COMMISSION

Arkansas Technical Reference Manual

Version 5.0

Approved in Docket
10-100-R

August 31, 2015

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

Arkansas Technical Reference Manual

Evaluation, Measurement & Verification Protocols

TRM Version 5.0 Volume 1: EM&V Protocols

Approved in Docket
10-100-R

August 31, 2015

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

Table of Contents

Section I: Introduction/Overview 1
A. Purpose of EM&V Protocols 2
B. Basics to Good Evaluation Practices 3
C. Role of Evaluation 4
D. Definitions of Process and Impact Evaluations 5
E. Evaluation Timing 9
F. Defining Program Baseline, Baseline Adjustments, and Data Collection Requirements 10
G. Defining Adjustment Factors 11
Section II: EM&V Protocols 16
PROTOCOL A: Program Tracking and Database Development 16
PROTOCOL B: Protocols for the Post-Implementation Verification of Energy Efficiency Program Measure Installations and Estimation of Energy and Demand Savings 19
PROTOCOL B1: Recommended Protocols for Gross Energy Evaluation 20
PROTOCOL B2: Recommended Protocols for Gross Demand Evaluation 20
PROTOCOL B3: Recommended Protocols for Participant Net Impact Evaluation 20
PROTOCOL B4: Sampling and Uncertainty Protocol 21
PROTOCOL C: Process Evaluation Guidance 22
PROTOCOL C1: Process Evaluation Structure and Timing 22
PROTOCOL C2: Process Evaluation Planning 26
PROTOCOL C3: Recommended Areas of Investigation in a Process Evaluation 27
PROTOCOL D: "Level of Effort" Protocols 30
PROTOCOL D1: Using Deemed Savings Values 31
PROTOCOL D2: M&V Protocols 31
PROTOCOL E: Protocols for Verification and Ongoing Modification of Deemed Savings Values 47
PROTOCOL E1: Revising and Updating Deemed Savings Values 47
PROTOCOL E2: IMplementation of Code Changes 48
PROTOCOL F: Protocols for the Determination of Net Program Impacts 49
PROTOCOL G: Provisions for Large Customers 57
PROTOCOL H: Technical Reference Manual (TRM) 60
PROTOCOL I: Role and Responsibilities of the Independent EM&V Monitor 66
PROTOCOL J: Behavior-Based Program Evaluation Protocol 68
PROTOCOL K: Leakage 82
References 88
Glossary 89
Index 96

List of Figures

Figure 1: Role of Program Evaluation in Program Implementation 5
Figure 2: Types of Data Collection Activities for Process and Impact Evaluations 8
Figure 3: Determining Timing for a Process Evaluation 24
Figure 4: Determining Need to Conduct a Process Evaluation 25
Figure 5: Decision Tree for Timing and Selection of NTG Research 55
Figure 6: Flowchart for Annual TRM Update Process 63
Figure 7: Detailed Process for Updating TRM 65
Figure 8: Randomized Controlled Trials with Opt-Out Enrollment 70
Figure 9: Example of Double-Counted Savings 76

List of Tables

Table 1: Comparison of Program Evaluation Objectives 6
Table 2: Summary of IPMVP Protocols 13
Table 3: Recommended Data Fields and Description 16
Table 4: Example of Data Legend for Database Tracking and Evaluation Purposes 17
Table 5: Example of Contractor Codes 18
Table 6: Suggested Data Collection Fields 18
Table 7: Annual Verification and TRM Update Timeline 64

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

Section I: Introduction/Overview

In September 2007, the Arkansas Public Service Commission (Commission) ordered each of the investor-owned, regulated utilities within the state to offer its customers energy savings programs. Now that Demand Side Management (DSM) programs are operating throughout Arkansas, it is important to periodically evaluate success factors and the level of market penetration for the programs to determine how effectively programs have been designed and implemented, estimate energy and demand savings attributable to the programs, and provide feedback and recommendations as programs mature. Program evaluation is not an afterthought; it is a vital part of successful program implementation and development. This document provides guidance and insight regarding some of the important Evaluation, Measurement & Verification (EM&V) methods and recommends best approaches to use specifically with the Arkansas DSM Program Portfolio.

This document is divided into two sections:

- Section I – Provides an overview of EM&V terms, methods, and approaches
- Section II – Provides the EM&V Protocols

The protocols presented in this Volume 1 of the TRM are based on a review of the current industry literature on a broad range of topics including:

- Master Plan and Project Management;
- Program Tracking and Database Development and Management;
- Protocols for Post-implementation Verification;
- Process Evaluation Guidance;
- "Level of Effort" or "Rigor" Protocols;
- Protocols for the addition of new measures and deemed savings values as well as the Verification and Ongoing Modification of Deemed Savings Values;
- Protocols for the Determination of Accurate Net Program Impacts (including net-to-gross estimates and non-energy benefits); and
- Provisions for Large Customers.

This review also draws heavily from the leading industry references used to guide EM&V activities for DSM Programs throughout the United States. This document also contains a complete bibliography of materials referenced to develop these EM&V Protocols for Arkansas, and are cited as appropriate. A Glossary of common EM&V terms is also included

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

A. PURPOSE OF EM&V PROTOCOLS

The purpose of these EM&V Protocols is to provide a common framework and set of reference points for conducting cost-effective DSM Program evaluations. These protocols describe the ***types*** of information that must be collected to conduct a comprehensive examination of a program's overall effectiveness, the recommended ***frequency*** for conducting these program evaluations, and the key ***metrics*** that must be reported during these evaluation activities.

These protocols provide additional guidance regarding the role of EM&V for DSM programs, as well as key definitions, recommendations regarding data capture and EM&V reporting formats. The EM&V plans developed for each organization or utility will describe the specific EM&V activities that will be completed during the evaluation cycle.

This document is not designed to address all of the key topics in the EM&V field. Rather, it is designed to address the specific topic areas in the Order 10-100-R.

Topics covered include:

- Evaluation Definitions
- Evaluation Planning
- The Value of Baseline Assessments
- The Role of Process Evaluations
- The Role of Impact Evaluations
- Specific Evaluation Protocols for the following topics:
    - Program Tracking and Database Development and Management;
    - Protocols for Post-implementation Verification;
    - Process Evaluation Guidance;
    - "Level of Effort" Protocols;
    - Protocols for Verification and Ongoing Modification of Deemed Savings Values;
    - Protocols for the Determination of Accurate Net Program Impacts (including net-to-gross estimates and non-energy benefits); and
    - Provisions for Large Customers

This document does not attempt to address the entire range of issues facing the current EM&V community. Therefore, these protocols will not address the following EM&V topics:

- Program Logic Models
- Market Transformation Studies
- Documenting Market Effects
- The Role of Economic Cost/Benefit Tests
- Monetizing Non-Energy Benefits

The goal of these EM&V Protocols is to provide the underlying evaluation framework for EM&V activities. The Arkansas Commission and the EM&V Parties Working Collaboratively (PWC) can then build on this framework to address additional EM&V issues as the energy efficiency market evolves and matures throughout the State. This document includes protocols on process evaluation and behavior-based energy efficiency programs. Subsequent revisions of this document should include protocols on a variety of other important energy efficiency topics as they reflect advances in the market. This will include the results of further analysis by the PWC regarding the inclusion of non-utility resources and other non-energy benefits that can be attributed to energy efficiency programs in the calculation of program cost-effectiveness.

B. BASICS TO GOOD EVALUATION PRACTICES

EM&V is the embodiment of the old adage "*Begin with the end in mind."* Ideally, program evaluation is built into the fabric of every program design so when the time comes for an evaluation the baseline data and measure data will be readily available and the evaluation activity will be a seamless, integrated process. However, the ideal is rarely the reality in program evaluations. Oftentimes, evaluators who are called upon to review program operations must rely on less-than-perfect data. The goal of these EM&V Protocols is to move the Arkansas EM&V activities closer to the ideal evaluation process by providing a clear description of what is needed and required.

The first step is a proper understanding of what evaluation means and entails. If a DSM program worked successfully, less energy will have been consumed by end-users during a specific time period. Evaluation is not measuring an event that happened; rather, it attempts to measure *"what did not happen. It is an estimate"* (Schiller Consulting 2010).

It is critical to understand that even the best evaluation practices are based on highly refined estimates, and do not provide absolute findings. The energy savings from a program cannot be determined directly from available energy use data but must be measured against what ***would have happened if the program did not exist***. Therefore, the goal of a good EM&V protocol is to identify the best approaches for determining reasonable and defensible estimates – largely based on surveys with program participants, non-participants and other market actors – about events that would have happened in the absence of the programs. With any evaluation, there is a level of ***risk*** that the estimations are inaccurate, and there are different points for acceptable margin of error (or levels of confidence). These protocols manage the risk of inaccuracy, and minimize the margin of error, by specifying the information and data points required to properly document savings and provide the best possible estimates of energy savings.

A second major issue regarding good EM&V practices relates to the level of effort required to obtain meaningful results, while managing evaluation costs. It is important to weigh the costs associated with obtaining additional, incremental information (or developing more precise estimates of program impacts, i.e., higher certainty), with the incremental costs associated with gathering and studying additional information.

Therefore, EM&V methodologies involve a series of tradeoffs guided by answering two questions:

Q1. "What is the comparison point?
Q2. "How good is good enough?"

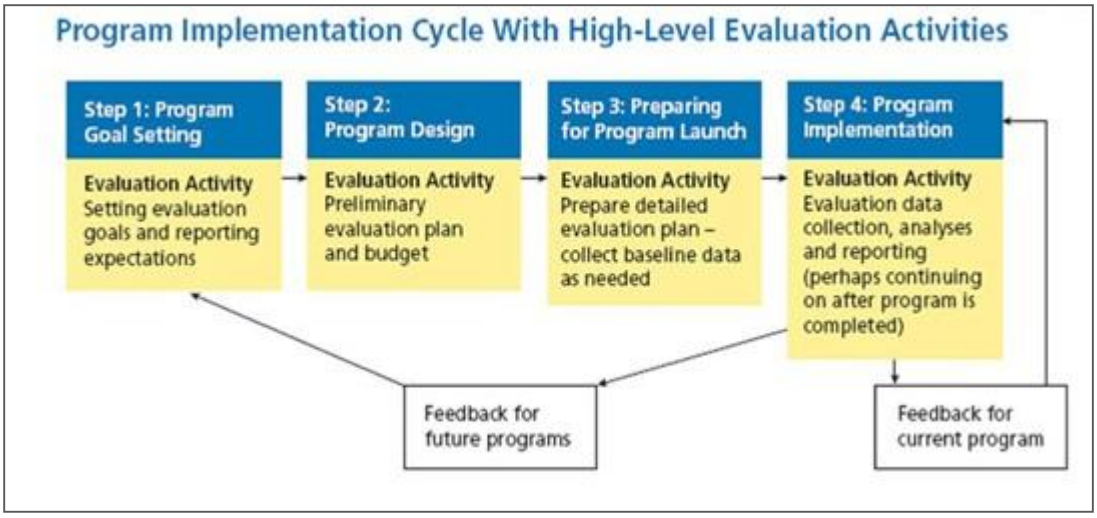
The answers to these questions are based on the size, scale, and scope of the overall program portfolio as it relates to the ultimate energy savings goals and objectives.

C. ROLE OF EVALUATION

The role of a Program Evaluation is to:

- **Quantify Results:** Document, measure and estimate the energy and demand savings of a program in order to determine how well it has achieved its goals and managed its budget.
- **Gain Understanding:** Determine why certain program effects occurred (or didn't occur) and identify ways to improve and refine current and future programs; also to help select future programs (NAPEE 2007).

The National Action Plan for Energy Efficiency (2007), which is heavily referenced in these protocols, provides an excellent visual representation of the role of program evaluation activities during the life cycle of a typical DSM program.



(Source: NAPEE 2007)

Figure 1: Role of Program Evaluation in Program Implementation

As Figure 1 shows, program evaluation is viewed as an ongoing process that provides information regarding changes in program direction and adjustments to program goals and objectives over time.

D. DEFINITIONS OF PROCESS AND IMPACT EVALUATIONS

The American Evaluation Association defines evaluation as "*assessing the strengths and weaknesses of programs, polices, personnel, products and organizations to improve their effectiveness."*

- **Process Evaluation** describes and assesses overall program structure, materials and activities.
- **Impact Evaluation** examines the long-term effects from a specific program, including unintended effects.
- **Process and Impact Evaluations** *work together* to provide a complete picture; activities related to these separate evaluation efforts often overlap.

According to the U.S. Department of Energy, program evaluations are used to make future decisions regarding program operations. The key questions that program evaluations answer include (EERE 2006):

- Should the program continue?
- How should the program be changed? Should it be expanded, consolidated, or replicated?
- Do program funds need to be reallocated?
- How can program operations be streamlined, refined or redesigned to better meet program objectives?
- What elements of the program can or should be discontinued?
- Is it time for this program to be discontinued?

The first step in a program evaluation is to specify the evaluation objectives. The two most common types of objectives are 1) studies that guide decisions about future program implementation and 2) estimates of quantified savings and insight regarding a program's overall performance. Table 1 compares these two types of evaluation activities.

Table 1: Comparison of Program Evaluation Objectives

Informational Needs	Evaluation Type
Efficiency of program implementation processes, e.g., to document the effectiveness of specific activities, what works and what does not work, where additional resources could be leveraged, participant satisfaction.	**Process**
Quantified outcomes that can be attributed to the program's activities, i.e., what are the results or outcomes that would not have occurred without the influence of the program. This is also called "net impacts."	**Impact**

(Source: Modified from EERE 2006)

Examples of common program evaluation objectives are:

- Assess the impact of the program on customer awareness and knowledge of energy efficiency actions.
- Measure customer response to "follow-up" program elements designed to encourage audit participants to implement recommendations.
- Examine program awareness, delivery channels, factors that influenced participation, program effects and customer satisfaction levels.
- Document energy efficiency actions taken by program participants compared to actions taken by non-participants.
- Estimate energy savings accruing from participation in the program over time; verify the reported energy savings as results of the program.
- Determine if there have been any changes in the building characteristics of program participants between program years.
- Evaluate the effectiveness of program modifications made in a specific fiscal year.
- Complete a customer segmentation analysis of the primary target population.
- Explore barriers to participation in program activities and develop recommendations for improving the promotion and targeting of existing services, as well as new program knowledge and services (Source: Expanded and modified from EERE 2006).

Process Evaluations are effective management tools that focus on improving both the design and delivery of energy efficiency programs. They are mostly commonly used to document program operations for new programs or those in a pilot or test mode. Since process evaluations examine program operations, they can identify ways to make program enhancements and improvements that reduce overall program costs, expedite program delivery, improve customer satisfaction, and fine-tune program objectives. These evaluations can also be used to assess the effectiveness of various incentive programs and rebated technologies. Process evaluations rely on a variety of qualitative and quantitative research methods, beginning with a review of program materials and records, conducting in-depth interviews with program staff and implementers, and

surveys with key customer and trade ally groups. Process evaluations can also provide feedback on ways to streamline and enhance data collection strategies for program operations (NAPEE 2007).

Impact Evaluations measure the change in energy usage and demand (kWh, kW, and therms) attributable to a program. They are based on a variety of approaches to quantify (estimate) energy savings, including statistical comparisons, engineering estimation, and modeling, metering, and billing analysis. The impact evaluation approach selected is primarily a function of the available budget, the technologies or energy end-use measures (EUMs) targeted in the program, the level of certainty of original program estimates, and the overall level of estimated savings attributable to the program (NAPEE 2007).

There are many decisions that affect the design of an impact evaluation. However, each impact evaluation should address the following seven major issues before the budget and evaluation plan are prepared:

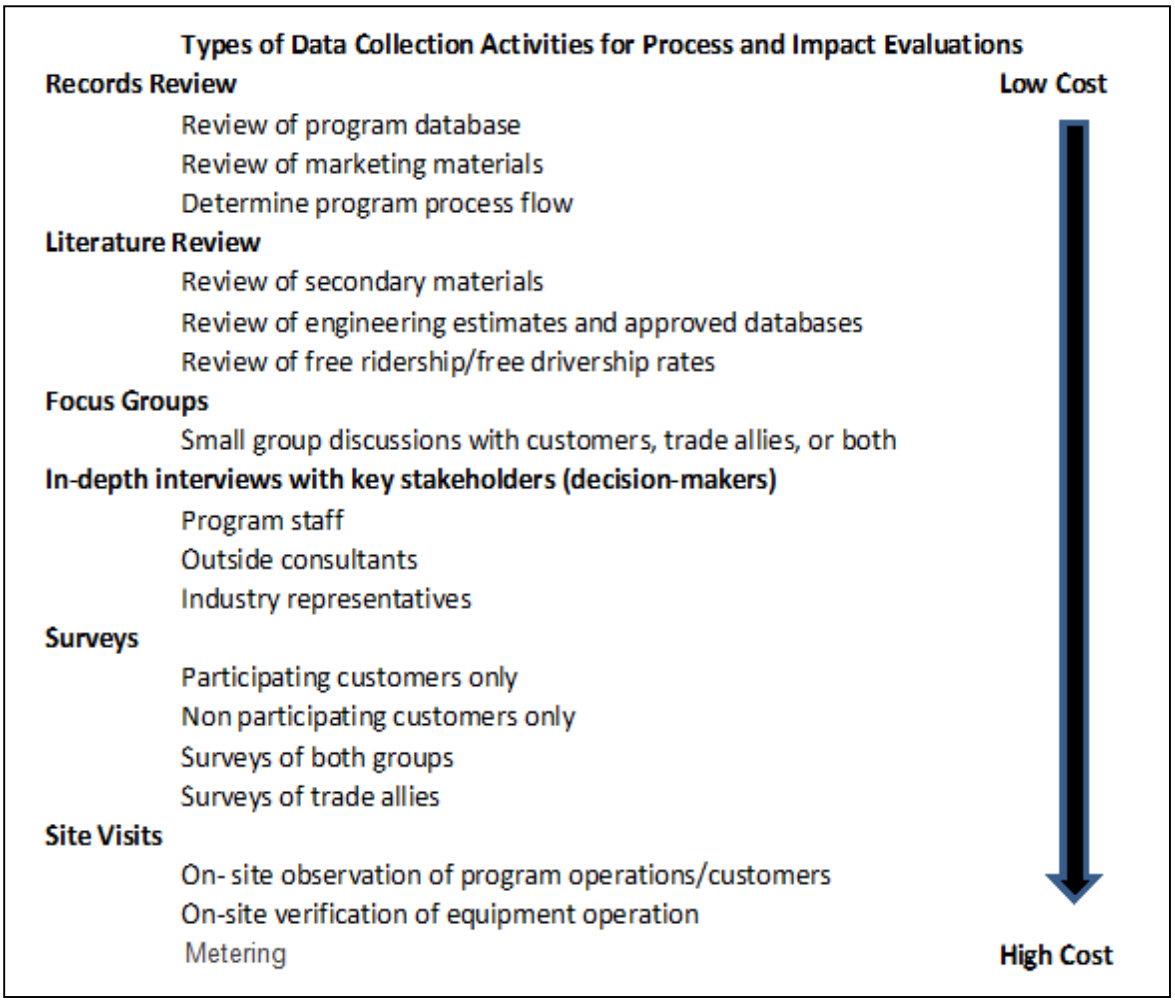
1. Define evaluation goals and scale (relative magnitude or comprehensiveness).
2. Set a time frame for evaluation and reporting expectations.
3. Set a spatial boundary for evaluation.
4. Define a program baseline, baseline adjustments, and data collection requirements.
5. Establish a budget in the context of information quality goals (level of acceptable margin of error and risk management).
6. Select impact evaluation approaches to estimate gross and net savings calculations.
7. Select who (or which type of organization) will conduct the evaluation (NAPEE 2007).

NAPEE 2007 summarizes what it considers to be the *"Best Practices in Evaluation."* These EM&V best practices are summarized below to provide further guidance for evaluation activities conducted in Arkansas.

- Incorporate an overall evaluation plan and budget into the program plan at the beginning of program planning.
- Adopt a more in-depth evaluation plan each program year.
- Prioritize evaluation resources where the risks are highest. This includes focusing impact evaluation activities on the most uncertain outcomes and highest potential savings. New and pilot programs have the most uncertain outcomes, as do newer technologies.
- Allow evaluation criteria to vary across program types to allow for education, outreach, and innovation.
- Conduct ongoing verification as part of the program process.
- Establish a program tracking system that includes necessary information for evaluation.
- Match evaluation techniques to the situation with regard to the evaluation costs, the level of precision required, and feasibility.

- Maintain separate staff for evaluation and for program implementation. Rely on an outside review of evaluations (e.g., state utility commission), especially if the program is implemented by internal utility staff. It is important that the program evaluation is an activity that is conducted independently of program operations.
- Evaluate regularly to refine programs as needed to meet changing market conditions.

Process and impact evaluations require data collection from a variety of sources. The timing, mix, and types of data collection activities must be specified in the EM&V plans for each program. Figure 2 presents the types of data collection activities for process and impact evaluations.



(Source: Reynolds, Johnson & Cullen 2008)

Figure 2: Types of Data Collection Activities for Process and Impact Evaluations

It is important to note that not every process or impact evaluation will require the complete set of data collection activities. Rather, the evaluation plan specifies the data collection strategies that will be used in each phase of the program evaluation as well as the anticipated budget expenditures for each data collection activity.

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

E. EVALUATION TIMING

All dates are approximate and subject to change based on commission rules and orders.

The decision regarding the appropriate evaluation time frame has two components:

1. **When and over what period of time the evaluation effort will take place.** A standard evaluation would begin before program implementation begins in order to collect important baseline data, and then continue for some time after the program is completed to analyze persistence of savings and other program elements. However, the actual timing of the evaluation is influenced by several factors, including:

    - What will be the time period of analyses (i.e., how many years)?
    - Will persistence of savings be determined, and if so, how?
    - What is the timing for policy decisions and evaluation planning?
    - What is the need for early feedback for program implementers?
    - Where is the program in its life cycle?
    - What are the evaluation data collection time lags?
    - What are the other regulatory and/or management oversight requirements to be addressed in this evaluation?
    - What information or data are needed to update specific energy and demand savings from the measure, and to quantify life estimates?
    - What is the timing and format required for the reporting process? Is a single, final program report needed, or are more frequent reports required?

    In general, program evaluations are conducted with a three-year plan. Process evaluations are usually conducted at the end of the first year of program operations and at the conclusion of the program period. Impact evaluations may be conducted annually or at the conclusion of Program Years 2 and 3, provided there are sufficient data to determine energy savings estimates and adjustments. The timing for the EM&V activities must be specified in the EM&V plans.

2. **What is the level of detail or "granularity" required for the evaluation analyses?** This relates to whether 15-minute, hourly, monthly, seasonal, or annual data collection and savings reporting are necessary. The granularity decision is based how the information will be used from the evaluation. Annual savings data provide an overview of program benefits. More detailed data are usually required for both cost-effectiveness analyses and resource planning purposes.

    If demand savings are to be calculated, the choice of definition (e.g., annual average, peak summer, coincident peak, etc.) is related to time granularity. When evaluating energy or demand savings, it is important to properly define the project boundaries (i.e., what equipment, systems, or facilities will be included in the analyses). Ideally, all primary effects (the intended savings) and secondary effects (unintended positive or negative effects), and all direct (at the project site) and indirect (at other sites) will be captured in the evaluation. The decision concerns whether savings will be evaluated for specific pieces of equipment. For example, the "boundary" may include motor savings or

light bulb savings estimates, the end-use system (e.g., the HVAC system or the lighting system), the entire facility or the entire energy supply and distribution system (Modified NAPEE 2007).

Therefore, the EM&V plans filed for the each program portfolio will stipulate the sampling strategy, the confidence and precision levels necessary to provide for a robust EM&V analysis of the savings estimates. Although the sampling strategy will vary by program and across the program portfolio, it must be fully described in each EM&V plan.

F. DEFINING PROGRAM BASELINE, BASELINE ADJUSTMENTS, AND DATA COLLECTION REQUIREMENTS

A major impact evaluation decision is defining the ***baseline conditions.*** The baseline reflects the conditions, including energy consumption, that were occurring before the launch of the program. Baseline definitions consist of site-specific issues and broader, policy-oriented considerations.

Site-specific issues include the characteristics of equipment in place before an efficiency measure is implemented as well as how and when the affected equipment or systems are operated. When defining the baseline, it is also important to consider where in the life cycle of the existing equipment or systems the new equipment was installed. The options are:

- *Early replacement* of equipment that had not reached the end of its useful life;
- *Failed equipment replacement,* with new energy efficient equipment installed; or
- *New construction*.

For each of these options, the two generic approaches to defining baselines are the *project-specific* and the *performance standard* procedure, described below.

Project-Specific Baseline

Under the project-specific procedure (used with all or a sample of the projects in a program), the baseline is defined by a specific technology or practice that would have been pursued, at the site of individual projects, if the program had not been implemented. For energy efficiency programs, the baseline is established by:

1. Assessing the existing equipment's energy consumption rate, based on measurements or historical data;
2. Completing an inventory of pre-retrofit equipment; or
3. Comparing to a control group's energy equipment (used where no standard exists or when the project is an "early replacement," i.e., implemented prior to equipment failure).

The most widely accepted method, and recommended for these EM&V Protocols is to define the baseline by determining what technologies the new equipment actually

replaces; that is, the baseline is related to actual historical base year energy consumption or demand, and carried forward to future years (NAPEE 2007).

Performance Standard Baseline

The Performance Standard Baseline approach avoids project-specific determinations of quantified energy and demand savings and instead develops a performance standard, which provides an estimate of baseline energy and demand for all the projects in a program. The assumption is that any project activity will produce *additional* savings if it has a "lower" baseline than the performance standard baseline. Performance standards are sometimes referred to as "multi-project baselines" because they can be used to estimate baseline savings for multiple project activities of the same type.

Under the performance standard procedure, baseline energy and demand are estimated by calculating an average (or better-than-average) consumption rate (or efficiency) for a blend of alternative technologies or practices. These standards are used in large-scale retrofit (early replacement) programs when the range of equipment being replaced and how it is operated cannot be individually determined. This would be the case, for example, in a residential compact fluorescent lamp (CFL) incentive program, where the types of lamps being replaced and the number of hours they operate cannot be determined for each home. Instead, studies are used to determine typical conditions. Another common use of performance standards is to define a baseline as the minimum efficiency standard for a piece of equipment as defined by a law, code, or standard industry practice. This is commonly used for new construction or equipment that replaces failed equipment (NAPEE 2007).

This approach is especially important when it is difficult to determine baselines, such as in new construction programs since no comparison period exists. However, the concepts of project and performance standard baseline definitions can still be used in these circumstances. The industry-accepted methods of defining new construction baselines are based on:

- The specifications of buildings that would have been built or equipment installed, without the influence of the program, at the specific site of each of construction project. This might be evaluated by standard practice evaluation or building plans and specifications that were prepared prior to the program being launched.
- Existing building codes and/or equipment standards.
- The performance of equipment, buildings, etc., in a comparison group of similar program non-participants.

G. DEFINING ADJUSTMENT FACTORS

A number of factors, or independent variables, affect energy and demand levels at the customer site and need to be considered when evaluating a DSM program. By accounting for these factors that are beyond the control of the program implementer or end-user, the adjustments bring energy use in the two time periods (before program launch and after or

during program delivery) to the same set of conditions. Common adjustment examples are:

- Weather corrections
- Changes in occupancy levels and hours
- Production levels
- Economic conditions
- Energy prices
- Changing codes/standards and common practice/changes to the baseline
- Interactions with other programs
- Changes in household or building characteristics.

These factors all affect total energy used and energy demand levels. There are a few methods for isolating the impacts of these factors in order to accurately attribute energy and demand reductions to the program being evaluated.

The decision to calculate gross energy savings or net energy savings depends on the program objectives and available evaluation resources. Gross savings are calculated when all that is needed is an estimate of the savings for each project participating in a program. The most common example of this is a project involving a contractor completing energy efficiency measures in a facility for the sole purpose of achieving energy savings (e.g., performance contracts). Net savings (which account for the range of independent variables discussed previously) are calculated when one wants to know the level of savings that occurred as a result of the program's influence on program participants and non-participants (Schiller Consulting 2010). The methods to adjust gross and net savings are summarized below:

Estimates of Gross Savings

Gross energy (or demand) savings are the expected change in energy consumption or demand that results directly from program-promoted actions (e.g., installing energy efficient lighting) taken by program participants under pre-defined assumed conditions.

Estimates of Net Savings

Net energy or demand savings refer to the portion of gross savings that is directly attributable to the influence of the program. This involves separating out the impacts that are a result of other influences, such as weather, energy prices, or even consumer self-motivation.

Most program evaluations seek estimates for both gross and net energy/demand savings. They require a net-to-gross estimate. Net-to-gross estimates refer to the ratio of total (gross) savings that are program-influenced (net savings). For example, a program's overall impact is determined by calculating the actual savings less the anticipated or projected estimates and then adjusted for free ridership and spillover.

$$\textbf{Impact} = \textbf{Actual post} - \textbf{Projected}_{\textbf{pre}} \pm \textbf{Adjustments}$$

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

The level of effort necessary to complete an evaluation is driven by the equipment type and data collection needs. The International Performance Measurement and Verification Protocol (IPMVP) is an important and widely used guidance document for determining the level of effort required to conduct EM&V studies. It provides guidelines about the "level of effort" required to document energy efficiency savings. The IPMVP presents various M&V options, summarized in Table 2, that help guide savings verification methods and levels of effort.

Table 2: Summary of IPMVP Protocols

IPMVP M&V Option	Measure Performance Characteristics	Data Requirements
Option A: Engineering calculations using spot or short-term measurements, and/or historical data	Constant performance	• Verified installation • Nameplate or stipulated performance parameters • Spot measurements • Run-time hour measurements
Option B: Engineering calculations using metered data.	Constant or variable performance	• Verified installation • Nameplate or stipulated performance parameters • End-use metered data
Option C: Analysis of utility meter (or sub-meter) data using techniques from simple comparison to multi-variate regression analysis.	Variable performance	• Verified installation • Utility metered or end-use metered data • Engineering estimate of savings input to SAE model
Option D: Calibrated energy simulation/modeling; calibrated with hourly or monthly utility billing data and/or end-use metering	Variable performance	• Verified installation • Spot measurements, run-time hour monitoring, and/or end-use metering to prepare inputs to models • Utility billing records, end-use metering, or other indices to calibrate models

(Source: IPMVP Protocols 2010)

Budget Considerations

Establishing a budget (i.e., funding level) for an evaluation requires consideration of all aspects of the evaluation process. The costs for high levels of confidence in the calculations must be balanced against the risks (and costs) associated with the value of savings being allocated to projects and programs. In this sense, evaluation processes involve some risk management decisions. Low-risk projects require less evaluation confidence and precision; high-risk projects require more confidence and precision. The acceptable level of uncertainty is often a subjective judgment based on the value of the energy and demand savings, the risk to the program associated with over- or underestimated savings, and a balance between encouraging efficiency actions and high levels of certainty. An important aspect of evaluation planning is deciding what level of risk is acceptable and thus determining the requirements for accuracy and a corresponding budget.

The level of acceptable risk is usually related to:

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

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

The determination of the appropriate sample size also affects the evaluation budget. Procedures such as a statistical power analysis help researchers determine the sample size needed to achieve the desired level of precision and confidence for key outcomes. In this way, researchers are assured of a statistically significant sample size.

The National Action Plan for Energy Efficiency (2007) suggests that a reasonable spending range for evaluation is 3 to 6 percent of program budgets. In general, on a unit-of-saved-energy basis, costs are inversely proportional to the magnitude of the savings (i.e., larger projects have lower per-unit evaluation costs) and directly proportional to uncertainty of predicted savings (i.e., projects with greater uncertainty in the predicted savings warrant higher EM&V costs, NAPEE 2007).

Cost-Effective Strategies for Program Evaluation

One effective way to minimize the EM&V evaluation costs – while maximizing the overall effectiveness of the final program evaluations – is to encourage (PWC) parties working collaboratively on EM&V activities where possible.

In particular, larger utilities should be encouraged to include smaller utilities in EM&V planning and research.[1] The potential cost savings from PWC on evaluation projects can be significant, especially for utilities with relatively small programs. Other advantages

[1] Although not mandatory, collaboration is a potentially cost-effective option that should be given consideration.

include a consistent methodological approach and a consistent reporting format. Similarly, market assessments and baseline studies naturally lend themselves to a regional approach because markets do not conform to service-territory boundaries.

Specific types of evaluation activities may be appropriate for collaboration:

1. Baseline Studies
2. Market Assessment Studies
3. Incremental Cost Studies
4. Process or Impact Evaluations where programs are substantially similar
5. DSM Potential Studies

If programs meet the following criteria, then conducting EM&V activities collectively via the PWC is strongly encouraged:

- **The programs delivered across multiple utility territories are very similar.** If the programs target the same customer class and offer the same measure mix, then there are significant cost savings to be gained by conducting a joint evaluation across the utility service territories. The most common DSM programs that encourage this PWC effort include residential CFL programs, residential weatherization programs, and energy audit programs.

- **The finding should be reported consistently across utility service territories where appropriate**. This PWC effort will only work if the participating utilities are tracking the customer data in the same manner across the joint programs. The individual utilities can report their EM&V activities separately but share the EM&V costs associated with executing these activities (Reynolds et al., 2009).

The Importance of Independence

EM&V requires third-party verification and reporting. Therefore, the organization selected to conduct any EM&V activities should be independent of the organizations involved in the program design, management, and implementation efforts. The evaluations should be conducted at an "arms-length distance," such that the verification professionals have no financial stake in the program or program components being evaluated beyond the contracted evaluation efforts.

Section II: EM&V Protocols

This section describes the recommended EM&V Protocols that should be incorporated in process and impact evaluations of the DSM programs pursuant to Docket No. 10-100-R.

PROTOCOL A: PROGRAM TRACKING AND DATABASE DEVELOPMENT

Protocol Scope: This protocol provides guidance to develop an effective DSM program tracking, evaluation and project database. It lists the key data elements that must be tracked, the key measure characteristics, key customer demographics and other data fields.

Customer Classes: All except self-directing customers

All tracking systems should capture all of the variables required to determine the energy savings. Please refer to the most recent version of the Deemed Savings estimates developed for the Arkansas Technical Reference Manual (TRM) as specified in the Deemed Savings No. 07-152-TF[2].

Table 3: Recommended Data Fields and Description

Recommended Data Fields	Description
Participating Customer Information • Unique customer identifier, such as account number • Customer contact information – name, mailing address, telephone number • Date/s of major customer milestone/s such as rebate application date, approval date, rebate processing date, etc.	Information used to readily identify customers for follow-up contact
Measure Specific Information • Measure Group (Equipment Type) • Equipment Fuel/Energy Source • Equipment size • Equipment quantity • Efficiency level • Estimated savings • Estimated incremental measure cost, if applicable • Equipment Useful Life • Measure Name - Text Description • Measure Code - Numerical Code* • Serial Number (where applicable) • Reported age of equipment replaced (if available) • Reported measure type of equipment replaced (if available) • Other inputs necessary for the use and compliance with the TRM	Information that documents the details of the equipment installed and equipment replaced under the program ***Measure Codes**: Ideally, all data should be captured in numeric format to facilitate data tracking and analysis. Therefore, a data legend should identify each measure type and contractor type. This data legend should also be clearly labeled in the program database's supporting materials.

[2] http://www.apscservices.info/EEInfo/TRM.pdf

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

Recommended Data Fields	Description
Vendor Specific Information • Name and Contact Information for Contractor • Contractor Type • Date of Installation • Cost of the installed equipment (if available) • Efficiency level of the installed equipment	To be collected when the measure is installed by a third-party vendor. This information can be determined from the supporting documentation provided to qualify for the program incentive.
Program Tracking Information • Date of the initial program contact/rebate information • Date of rebate/incentive paid • Incentive amount paid to date • Incentive amounts remaining • Application Status (i.e., number of applications approved, pending or denied) • Reason and Reason code for application denial	Information to determine program cost effectiveness and timing for rebate applications and processing
Program Costs • Overall program budgets • Program expenditures to date • Incentive Costs • Administrative Costs • Marketing/Outreach Costs • Evaluation Costs	This information is directly related to program expenses. This information may be tracked in a separate worksheet from measure costs; however, the totals should be reported annually.
Marketing and Outreach Activities • Advertising and marketing spending levels • Media schedules • Summary of number of community events/outreach activities • Other media activities – estimated impressions via mailings, television/radio, print ads	The program implementers should be able to provide separate documentation regarding the type, number, and estimated impressions made for each marketing or outreach activity.

Table 4: Example of Data Legend for Database Tracking and Evaluation Purposes

Example Measure Category	Example Measure Code
Air Source Heat Pump	1
Room Air Conditioner	2
Central Air Conditioner	3
Natural Gas Furnace	4
Storage Water Heater (Gas)	5
Tankless Water Heater (Gas)	6
Storage Water Heater (Electric)	7
Heat Pump Water Heater	8
Attic Insulation	9
Wall Insulation	10

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

Similarly, the contractor type could also be identified by a category and a numeric code to facilitate analysis and tracking. Ideally, the program database and tracking system would be linked to the utility's or energy provider's current Customer Information System so that it can be updated regularly to verify eligibility.

Table 5: Example of Contractor Codes

Example Contractor Type	Example Contractor Code
Architect	11
Engineer	22
Plumber	33
HVAC	44
Insulation Installer	55
Home Builder (Production)	66
Home Builder (Custom)	67
Specialty	90

"Best practices" regarding database tracking and development also suggest capturing additional types of information during data collection to facilitate EM&V. Examples are provided below in Table 6.

Table 6: Suggested Data Collection Fields

Suggested Data Collection Fields	Description
• **Premise Characteristics** • Housing Type • Number of Occupants • Estimated/Actual Square Footage	This information includes descriptions of the housing type and similar data points asked of participants during the measure installation.
• **Measure Characteristics** • Efficiency level of equipment removed (retrofit only) • Model level for equipment removed (retrofit only)	This information is commonly captured by the contractor or recorded from the invoice and could be tracked in the program database.

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

PROTOCOL B: PROTOCOLS FOR THE POST-IMPLEMENTATION VERIFICATION OF ENERGY EFFICIENCY PROGRAM MEASURE INSTALLATIONS AND ESTIMATION OF ENERGY AND DEMAND SAVINGS

Protocol Scope: This protocol specifies the types and categories of measures that require post-implementation verification, the recommended timing for these activities, and the key data to capture during on-site inspections. For more detailed information regarding the data collection requirements for on-site inspections refer to the IPMVP listed on www.evo-world.org and the requirements listed in the Deemed Savings Docket 07-152-TF.

Customer Classes: All except self-directing customers

Calculating Energy and Demand Savings: For efficiency programs, determining energy savings is the most common goal of impact evaluations. Energy usage and savings are expressed in terms of consumption over a set time period and defined in well understood terms (e.g., therms of natural gas consumed per month, megawatt hours [MWh] of electricity consumed over a year, season, or month). Energy savings results may also be reported by costing period, which break the year into several periods coinciding with a utility rate schedule.

Special cases of savings by period include demand savings for peak and off-peak periods or summer and winter periods.

Examples of demand savings definitions are:

Annual average demand (MW) savings – Total annual energy (MWh) savings divided by the hours in the year (8,760).

Peak demand reductions – The maximum amount of demand reduction achieved during a period of time. This time period should be clearly defined, whether it is annual, seasonal, or during a specific period of time, such as summer weekday afternoons or winter peak billing hours.

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 what portion of a utility's system 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).

PROTOCOL B1: RECOMMENDED PROTOCOLS FOR GROSS ENERGY EVALUATION
1. A Simple Engineering Model (SEM) with M&V equal to IPMVP Option A and meeting all requirements for this M&V Protocol is recommended. Other approaches and options should be considered as evaluation requirements or studies dictate.
2. Normalized Annual Consumption (NAC) using pre- and post-program participation consumption from utility bills from the appropriate meters related to the measures installed, weather-normalized, using the identified weather data to normalize for heating and/or cooling as appropriate. Twelve or more months' pre-retrofit and twelve months' post-retrofit consumption data is recommended.

(Source: CA Evaluators' Protocols 2006)

PROTOCOL B2: RECOMMENDED PROTOCOLS FOR GROSS DEMAND EVALUATION
1. Requires using secondary data to estimate demand impacts as a function of energy savings. End-use savings load shapes or end-use load shapes will be used to estimate demand impacts as available:
a. End-use savings load shapes, end-use load shapes or allocation factors from simulations conducted for Arkansas TRM as available.
b. Allocation factors from forecasting models or utility forecasting models through the evaluation plan review process such as econometric, end use, load forecast and other models as appropriate.
c. Allocation based on end-use savings load shapes or end-use load shapes from other studies for related programs/similar markets with approval through the evaluation review process as applicable.

(Source: Modified from the CA Evaluators' Protocols 2006)

PROTOCOL B3: RECOMMENDED PROTOCOLS FOR PARTICIPANT NET IMPACT EVALUATION
1. Analysis of utility consumption data that addresses the issue of self-selection bias for both participants and non-participants.
2. Enhanced self-report method using other data sources relevant to the decision to install/adopt. These could include: a record/business policy and paper review, examination of similar decisions, interviews with multiple actors at end-user site including participants and non-participants, interviews with mid-stream and upstream market actors, reviews of standard buildings and equipment installation practices by builders and/or stocking practices.

(Source: Modified from the CA Evaluators' Protocols 2006)

PROTOCOL B4: SAMPLING AND UNCERTAINTY PROTOCOL
Level Gross Impact Options: **Simplified Engineering Models (SEM)**: The sampling unit is the premise or site where the measure was installed. The sample size selected must be justified in the evaluation plan and approved as part of the evaluation planning process.
Protocols for Estimating Net Impacts: If the method used for estimating net energy and demand impacts and net-to-gross ratios (NTGR) is regression-based; there are no relative precision targets. For both impacts and NTGR calculation, evaluators are expected to conduct, at a minimum, a statistical power analysis as a way of initially estimating the required sample size. Other information can be taken into account such as professional judgment and prior evaluations of similar programs or similar measures.
Normalized Annual Consumption (NAC) Models: If NAC models are used, there are no targets for relative precision. This is due to the fact that NAC models are typically estimated for all participants for whom there is an adequate amount of pre- and post-billing data. Thus, there is no sampling error. However, if sampling is conducted, either a power analysis or justification based upon prior evaluations of similar programs must be used to determine sample sizes. The sample size selected should be justified in the evaluation plan and approved as part of the evaluation planning process.

(Source: Modified from the CA Evaluators' Protocols 2006)

PROTOCOL C: PROCESS EVALUATION GUIDANCE

Protocol Scope: This protocol provides guidance regarding the timing and scope for process evaluations of the Arkansas utility programs. Process evaluations focus on determining the overall effectiveness of program delivery, identifying opportunities for program improvements and assessing key program metrics, including participation rates, market barriers, and overall program operations.

PROTOCOL C1: PROCESS EVALUATION STRUCTURE AND TIMING

Protocol Scope: This protocol section provides additional guidance on how to best structure process evaluations at the state, portfolio, program, service, and market sector level. Process evaluations need to be structured to meet the specific goals and objectives at a particular point in time.

Customer Classes: All except self-directing customers

Program Types: All

Approach: The process evaluation decision-maker, either the utility or third-party administrator, should determine if a process evaluation is needed based on any of the criteria described in Protocols C1 and C2, which summarize the two major criteria for determining if a process evaluation is necessary. The first criterion is to determine if it is time for a process evaluation; the second criterion is to determine if there is a need for a process evaluation. Figures 3 and 4 illustrate this decision-making process.

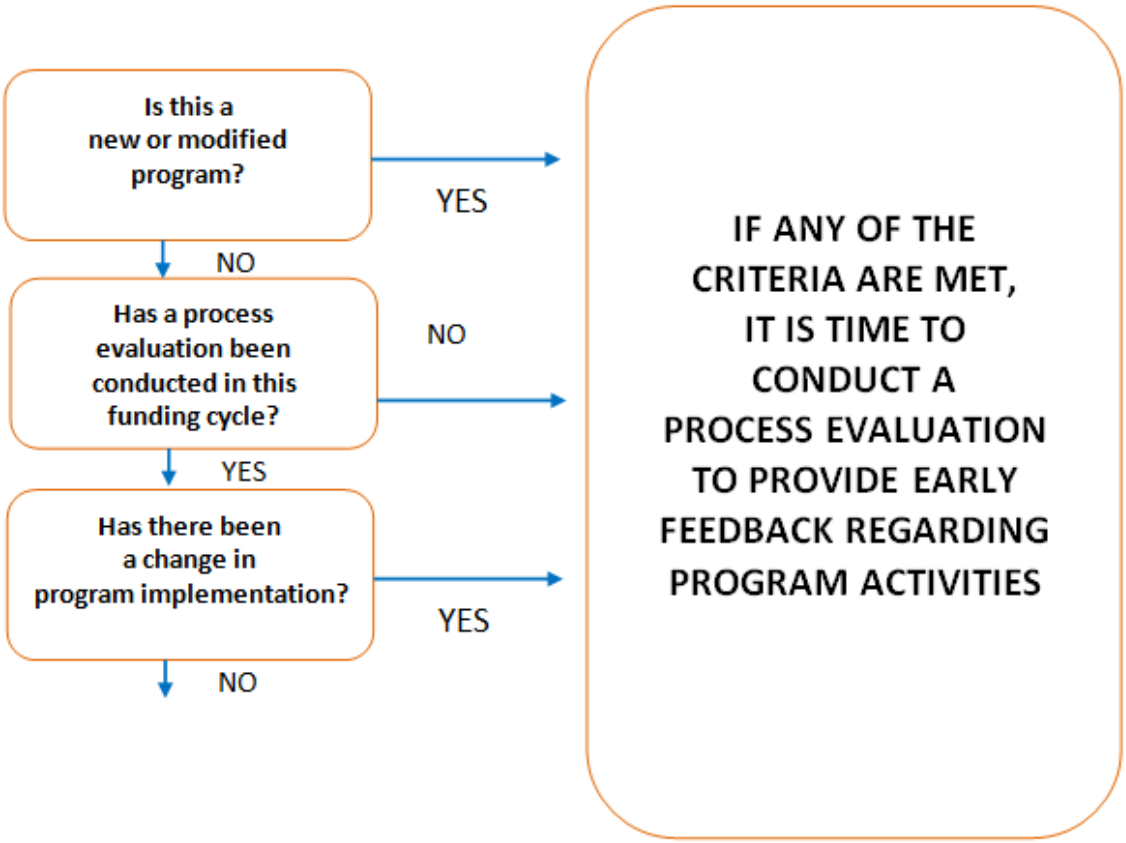
Keywords: "timing; portfolio level evaluations; process evaluation structure; diagnostic process evaluations; under-performing programs; programs not meeting targets"

PROTOCOL C1: DETERMINING APPROPRIATE TIMING TO CONDUCT A PROCESS EVALUATION
1. **New and Innovative Components:** If the program has new or innovative components that have not been evaluated previously, then a process evaluation needs to be included in the overall evaluation plan for assessing their level of success in the current program and their applicability for use in other programs.
2. **No Previous Process Evaluation:** If the program has not had a comprehensive process evaluation during the previous funding cycle, then the Program Administrator should consider including a process evaluation in the evaluation plan.
3. **New Vendor or Contractor**: If the program is a continuing or ongoing program, but is now being implemented, in whole or in part, by a different vendor than in the previous program cycle, then the administrator should consider including a process evaluation in the evaluation plan to determine if the new vendor is effectively implementing the program.
If any of these criteria are met, it is time to conduct a process evaluation. **If none of these criteria are met, then the evaluation decision-maker should proceed to Step 2 in the Process Evaluation Decision Map.**

PROTOCOL C1: DETERMINING APPROPRIATE CONDITIONS TO CONDUCT A PROCESS EVALUATION
Process evaluations may also be needed to diagnose areas where the program is not performing as expected. These conditions may include the following:
1. Impact Problems: Are program impacts lower or slower than expected?
2. Informational/Educational Objectives: Are the educational or informational goals not meeting program goals?
3. Participation Problems: Are the participation rates lower or slower than expected?
4. Operational Challenges: Are the program's operational or management structure slow to get up and running or not meeting program administrative needs?
5. Cost-Effectiveness: Is the program's cost-effectiveness less than expected?
6. Negative Feedback: Do participants report problems with the program or low rates of satisfaction?
7. Market Effects: Is the program producing the intended market effects?
If any of the criteria is met, a process evaluation is needed to identify ways to address and correct these operational issues. **If none of these criteria is met in either Step 1 or Step 2, then a process evaluation is not needed at this time.** **Re-evaluate the need for a process evaluation at the end of the program year.**

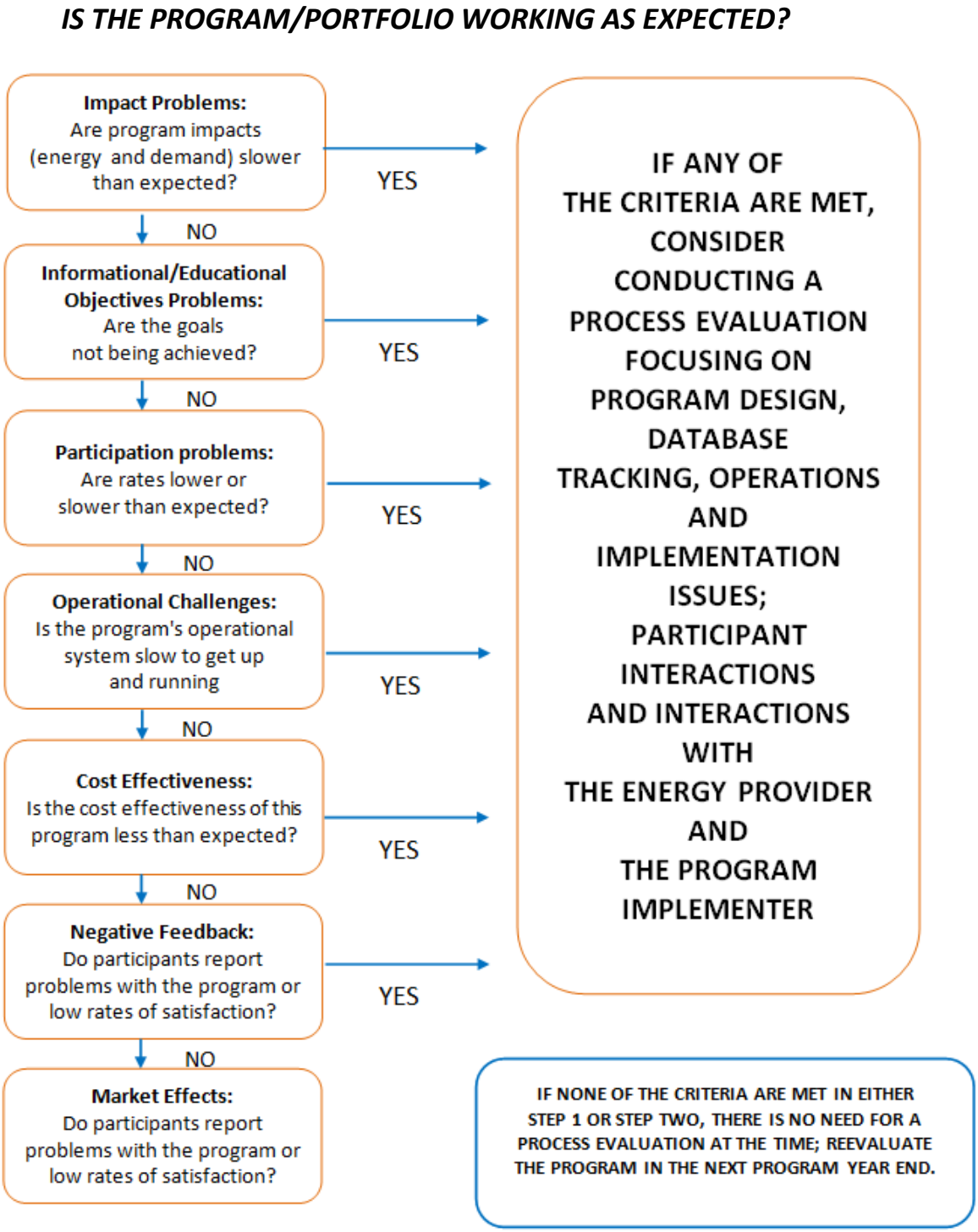
APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

IS IT TIME FOR A PROCESS EVALUATION?



(Source: Johnson & Eisenberg 2011, p. 21.)

Figure 3: Determining Timing for a Process Evaluation



(Source: Modified from Johnson & Eisenberg 2011, p. 22)

Figure 4: Determining Need to Conduct a Process Evaluation

Additional Guidance for Conducting Limited/Focused Process Evaluations

In all cases, the evaluator should conduct a limited or focused process evaluation consisting of a review of the program database and staff interviews to determine each program's progress throughout the evaluation cycle. The findings from these activities will serve to:

- Provide a progress report for each recommendation for program improvement made in previously conducted evaluations. For each evaluation recommendation, the report should indicate whether the recommendation has been accepted and implemented, rejected, or is still under consideration. If the recommendation is rejected, an explanation of the reason for rejection should be provided. If a recommendation is still under consideration, then an explanation should be provided for the steps underway to reach an implementation decision for that recommendation;
- Identify the progress made towards achieving the objectives as described in the Commission Checklist; and
- Identify any issues that may need to be explored more fully in future program evaluations.

PROTOCOL C2: PROCESS EVALUATION PLANNING

Protocol Scope: This protocol provides guidance on the key issues that should be addressed in process evaluations. It is especially important to focus on the aspects of program operations to address any deficiencies identified in the Process Evaluation Decision Map, Figure 4.

Customer Classes: All

Program Types: All

Approach: The process evaluation plan should use the following outline to identify the key researchable issues that must be addressed in the process evaluation. This outline applies to process evaluations conducted at the program, portfolio, and state level.

Keywords: "process evaluation planning; EM&V plan process evaluation timing; portfolio level process evaluations; process evaluation structure; process evaluation components; process evaluation scope"

PROTOCOL C2: RECOMMENDED ELEMENTS OF A PROCESS EVALUATION PLAN
Introduction: Description of the program or portfolio under investigation; specific characteristics of the energy organization providing the program including current marketing, educational or outreach activities and delivery channels
Process Evaluation Methodology: Process evaluation objectives, researchable issues, and a description of how specific evaluation tactics will address the key researchable issues including the proposed sampling methodology for program/third-party staff, key stakeholders, trade allies/vendors, and customers. The sampling methodology should be clearly explained with specific targets of completed surveys or interviews clearly described in the EM&V Plan.
Timeline: Summarized by key tasks identifying the length of the process evaluation and key dates for completion of major milestones
Budget: Costs of conducting the process evaluation by specific tasks and deliverables

(Source: Modified and Expanded from the California Evaluators' Protocols - TechMarket Works 2006).

While Protocol C2 provides a general outline of the key elements that should be included in a process evaluation plan, Protocol C3 provides more detailed information regarding the key areas for investigation that need to be addressed in a process evaluation. Protocol C3 also identifies those areas that are most applicable to new programs or pilot programs, those areas that should be investigated when the program is experiencing specific operational issues or challenges, and those topic areas that should be covered in all process evaluations.

PROTOCOL C3: RECOMMENDED AREAS OF INVESTIGATION IN A PROCESS EVALUATION	
Program Design	**Additional Guidance**
• Program design and design characteristics, and program design process	This area is especially important to address in first-year evaluations and evaluations of pilot programs.
• The program mission, vision and goal setting and goal setting process	
• Assessment or development of program and market operations theories	
• Use of new or best practices	
Program Administration	**Additional Guidance**
• The program management process	This area should be covered in all process evaluations, but it is especially important to address in those evaluations where operational or administrative deficiencies exist.
• Program staffing allocation and requirements	
• Management and staff skill and training needs	
• Program tracking information and information support systems	
• Reporting and the relationship between effective tracking and management, including operational and financial management	
Program Implementation and Delivery	**Additional Guidance**
• Description and assessment of the program implementation and delivery process	This is critical to gathering the information necessary to assess the program's operational flow.

PROTOCOL C3: RECOMMENDED AREAS OF INVESTIGATION IN A PROCESS EVALUATION	
• Program marketing, outreaching, and targeting activities	These are areas that should be addressed if the program is not meeting its participation goals or if the program is under-performing.
• Quality control methods or operational issues	
• Program management and management's operational practices	
• Program delivery systems, components and implementation practices	
• Program targeting, marketing, and outreach efforts	The process evaluator should request copies of all marketing and outreach materials and include an assessment as part of the document review task.
• Program goal attainment and goal-associated implementation processes and results	These areas should be addressed in all process evaluations, but are especially important if the program is under-performing regarding savings or participation rates.
• Program timing, timelines and time sensitive accomplishments	
• Quality control procedures and processes	
• Documentation of program tracking methods and reporting formats	This is a key element of the review of the program database and the evaluator should request copies of the program records or extracts along with the data dictionary.
• Customer interaction and satisfaction (both overall satisfaction and satisfaction with key program components, including satisfaction with key customer-product-provider relationships and support services)	These topics should be investigated in the customer surveys and should be a priority if the program is experiencing negative feedback or lower than expected participation rates or energy savings.
• Customer or participant's energy efficiency or load reduction needs and the ability of the program to deliver on those needs	
• Market allies interaction and satisfaction with the program	
• Reasons for a low level of market effects and spillover	
• Intended or unanticipated market effects	

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

The process evaluation report should include the following reporting requirements:

1. Detailed Program Description. The process evaluation report should present a detailed operational description of the program that focuses on the program components being evaluated. The use of a program flow model is highly recommended. The report should provide sufficient detail so that readers are able to understand program operations and the likely results of the recommended program changes.

2. Program Theory. The process evaluation should include a presentation of the program theory. If the program theory is not available, or cannot be provided in time for the evaluation report due date, the evaluator should include a summary program theory built from the evaluation team's program knowledge. It should be complete enough for the reader to understand the context for program recommendations, but does not need to be a finely detailed program theory or logic model.

3. Support for Recommended Program Changes. All recommendations need to be adequately supported. Each recommendation should be included in the Executive Summary and then presented in the Findings text along with the analysis conducted and the theoretical basis for making the recommendation. The Findings section should also include a description of how the recommendation is expected to help the program, including the expected effect that implementing the change will have on the operations of the program.

4. Detailed Presentation of Findings. A detailed presentation of the findings from the study is essential. The presentation should convey the conditions of the program being evaluated and should provide enough detail so that any reader can understand the findings and the implications of the overall operations of the program and its cost-effectiveness (Modified from the CA Evaluators' Protocols 2006).

PROTOCOL D: "LEVEL OF EFFORT" PROTOCOLS

Protocol Scope: This section addresses appropriate levels of effort for program evaluation activities based on measure type. This protocol specifies the following: 1) which measures are best suited to relying on deemed savings estimates; 2) which measures are best served through simplified EM&V activities such as on-site inspections or an engineering review; and 3) which measures require full EM&V activities, as defined by the IPMVP. These determinations are based on the measure characteristics, usage patterns, and program types. They are often conducted jointly with a desk review of all related measure-supporting documents, such as invoices, technical studies, and energy audits. In many cases, lower levels of effort are sufficient for estimating measure impact, which may help increase the overall cost-effectiveness of EM&V activities.

Customer Classes: All except self-directing customers

Protocol D1 specifically refers to those measures that have been defined in the Arkansas TRM as "deemed savings measures." Relying on agreed upon energy savings using the appropriately updated deemed savings values described in the Arkansas TRM is a valid approach for prescriptive and direct-install measures used in a traditional manner including, but not limited to: water heaters, furnaces, boilers, food service equipment, low-flow showerheads and faucet aerators. Please consult this document for the most up-to-date listing of applicable measures filed in Deemed Savings Docket 07-152-TF. This document includes the appropriate deemed savings measures for electric and gas measures in the residential, commercial and industrial (C&I) markets.

Measure Lives and Failure Rates

Measure lives listed in this TRM should be used, unless a reason is given for deviating from a measure life, along with support for the deviation.

- "Failure" is defined as an instance where an implementation contractor reports that a measure has been installed, but a subsequent inspection finds that the equipment is non-operational and/or not properly installed and that difference has not been accounted for elsewhere.
- "Failure rate" should be defined as the percent of inspected installation sites where any equipment fails inspection (i.e., the equipment is either not installed or not operating) and that possibility has not been otherwise accounted for.

Note, the definition of failure is intended to not count issues related to persistence or normal measure lives.

PROTOCOL D1: USING DEEMED SAVINGS VALUES

Deemed savings approaches are most commonly used for measures that involve simple new construction or retrofit energy efficiency measures with well-defined applications[3]. The deemed savings approach is most applicable when most of the following conditions are true:

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

Deemed values should be based on reliable, traceable, and documented sources of information, with an emphasis on rigorous evaluations of similar measures or programs in other jurisdictions. Other appropriate secondary sources of deemed parameters may include:

a. Standard tables or algorithms from recognized sources that indicate 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).
b. Manufacturer's specifications
c. Building occupancy schedules
d. Maintenance logs

Deemed saving values may vary depending on whether the measure was removed prior to the end of its expected useful life or replaced on burnout. In the case of replacement prior to the expected useful life, extra care should be taken that the savings are not over estimated.

(Source: Modified and Expanded from the CA Evaluators' Protocols 2006)

PROTOCOL D2: M&V PROTOCOLS

The IPMVP protocols require two elements of measurement and verification (M&V).

1. Verification that the measure or project has the potential to perform (is installed and operating correctly); and
2. Verification that energy savings is occurring.

   An M&V approach should be pursued only if the value of the reduction in uncertainty it yields exceeds its cost. Uncertainty in a savings estimate is partly a function of the variability in the energy use from one application to the next. To address this, IPMVP classifies projects as having high or low *energy variation* and high or low *value*. Consult the full IPMVP M&V Plan requirements to ensure that the EM&V activities conform to the specific data collection requirements. The four options presented in the IPMVP for savings estimation are listed next.

(Source: Schiller Consulting 2010 and IPMVP Protocol)

[3] Examples include T-8 lighting retrofits in office buildings or compact CFL giveaways for residential utility customers. In each example, an assumption is made about the average wattage savings and the average hours of operation combined with the effort of verifying that the T-8s were installed and the CFLs were actually provided to residents.

The following set of measure-specific protocols has been adapted and modified from the Northeast Energy Efficiency Partnerships (NEEP) EM&V Protocols (2010), which have combined current "best practices" for determining program estimates with both a recommended approach and an alternative approach consistent with the IPMVP approach. These protocols have been modified to reflect characteristics of the current Arkansas DSM programs. These protocols are measure-specific, but are not meant to be exhaustive or exclusive. Please consult the full list of all deemed measures in the Arkansas TRM for a full listing of appropriate deemed measures.

Option A– Retrofit Isolation: Key Parameter Measurement
Typical Application: Lighting retrofit where power draw is key performance parameter. This is measured on a constant and periodic basis.
Savings Determined by: Estimating lighting operating hours based on building schedules and occupant behavior.
Option B – Retrofit Isolation: All Parameter Measurement
Typical Application: Variable-speed drive and controls installed on a motor to adjust pump flow; there are variable/differing operating characteristics
Savings Determined by: *Spot Metering or collection of interval data.* An example is measuring electric power with a kW meter installed on electrical supply to the motor, which reads power at specified intervals. In baseline period this metering may be in place for a period of time, as appropriate to verify motor loading levels. The meter is in place throughout reporting period to track variations in power use.
Option C – Whole Facility Energy Bill Analysis
Typical Application: Multi-faceted energy management program affecting many systems in a facility.
Savings Determined by: Measuring energy use with the gas and electric utility meters for a 12-month baseline period and throughout the reporting period.
Option D – Whole Facility Calibrated Simulation
Typical Applications: Multifaceted energy management program affecting many systems in a facility but where no meter existed in the baseline period – such as new construction.
Savings Determined by: Energy use measurements, after installation of gas and electric meters that are used to calibrate a simulation. Baseline energy use, determined using calibrated simulation, is compared to either a simulation of reporting period energy use or actual meter data.

RESIDENTIAL MEASURES

Residential measures covered here include:

- Residential Lighting
- Residential Central Air Conditioning
- Residential Comprehensive Multi-Measure Retrofit
- Residential Natural Gas Boilers And Furnaces

RESIDENTIAL LIGHTING **Recommended EM&V Methods**		
This category is limited to single-family residential lighting exclusive of specialty hard-to-reach and multifamily programs. These measures include new construction, retrofit, direct install, and retail lighting programs.		
Characteristic	**Approach**	**Additional Comments**
Program Tracking	• Initial gross energy and demand savings, initial net impacts as applicable. • Baseline quantity and wattage, installed quantity and wattage, location (as available), hours of use, in-service rate, HVAC interaction.	Additional parameters useful for quality control and also for evaluation design, e.g. sampling. The tracking needs to conform with the requirements specified in Protocol A.
Recommended M&V Method	On-site inspections with partial measurements on a sample of program participants (Option A). Complete "socket counts" by room and fixture type provide key data for impact evaluations, baseline studies, and hours-of-use studies. Questions on purchasing habits and "shelf" stock inform in-service rate research. Site visits with time-of-use lighting loggers are the most defensible approach to residential lighting programs. Collection of basic heating and cooling system information can be helpful in assessing interactive savings effects.	Time-of-use lighting loggers on a sample of lamps and fixtures, typically by room type.
Acceptable Alternative M&V Methods	Alternatively, the Verification component can rely upon customer telephone surveys to obtain information such as socket counts, hours of use, and purchasing habits. These findings can be supplemented with literature reviews of other lighting studies to determine best estimates. This type of verification is an acceptable degree for rigor for these types of program installations.	The details regarding the alternative M&V approach must be documented in the Evaluation Plan submitted to the PSC. The EM&V for residential lighting may also address the issue of changing baselines due to the EISA phase out of standard incandescent lamps.

(Source: Modified from the NEEP EM&V Protocols 2010)

RESIDENTIAL CENTRAL AIR CONDITIONING **Summary of Recommended EM&V Methods**		
This category is limited to central air conditioning (CAC) installed as a stand-alone measure and excludes CAC installed through comprehensive new construction programs. This category does not include ENERGY STAR room air conditioners or other "space cooling" measures.		
Characteristic	**Approach**	**Additional Comments**
Program Tracking	• Initial gross energy and demand savings, as well as initial net impacts as applicable. • Number of installed units, unit capacity, baseline and installed efficiency, and full load cooling hours.	Additional parameters useful for quality control and also for evaluation design as specified in an EM&V plan.
Recommended M&V Method	On-site inspections with partial measurements on a sample of program participants (Option A). Site visits with short-term metering offers the most defensible approach to residential CAC programs.	Metering methods may include time-of-use loggers and spot power measurements. Logging load and energy draw data are recommended
Alternative M&V Methods	On-site inspections with metering that fully isolates the entire CAC system (Option B) is an acceptable alternative approach.	Metering would be interval kW measurements on both the outdoor compressor and indoor fan units.
	Billing analysis (Option C) is a reasonable energy evaluation method for residential CAC at lower cost.	Billing analysis alone generally cannot quantify demand impacts.
	Calibrated simulation modeling (Option D) is a viable alternative and can be effective at capturing measure interaction. CAC simulation modeling may be appropriate for evaluating comprehensive cooling measures.	Metering can be used to calibrate the model. Such metering may include whole premise interval kW recorders with some end use metering.

(Source: Modified from the NEEP EM&V Protocols 2010)

RESIDENTIAL COMPREHENSIVE MULTI-MEASURE RETROFIT **Summary of Recommended EM&V Methods**		
This category encompasses comprehensive multi-measure retrofit installations in residential homes. Sometimes called "deep retrofits" or "home energy services," these measures are characterized by a whole-home approach that typically involves an audit followed by efficiency recommendations for multiple end uses and technologies. The comprehensive residential approach tends to be electric-centric but may also span fuel measures such as water heating, boilers or furnaces.		
Characteristic	**Approach**	**Additional Comments**
Program Tracking	• Initial estimates of gross energy and demand savings, as well as initial net impacts as applicable. • Detail on individual measures, such as: air conditioner, heat pump, boiler/furnace, water heater quantities and sizes; baseline and installed equipment efficiencies; home square footage; insulation and weatherization actions.	Additional parameters useful for quality control and also for evaluation design, e.g. sampling as specified in the EM&V plan.
Recommended M&V Method	On-site inspections with partial measurements on a sample of program participants (Option A). Site visits with visual inspections, quality of installation assessments, interviews, and short-term metering for selected electric measures. Simple engineering models of savings impacts.	Metering is limited to time-of-use loggers on lighting and HVAC equipment supported by spot power measurements.
Alternative M&V Methods	For measures that save both natural gas and electricity, an option is to pair the Option A approach with a billing analysis (Option C) to determine gas impacts. Diagnostic testing of HVAC equipment, blower door, and duct blaster tests adds rigor and reduces uncertainty to savings estimates for envelope measures.	Evaluators should design and evaluation plan to achieve the identified objectives of the EM&V activities
	Calibrated simulation modeling (Option D) is a viable alternative and it is capable of capturing measure interaction. This approach may be most appropriate for comprehensive multi- measures.	Metering can be used to document HVAC system and whole premise interval kW and recorders with some temperature measurements.

(Source: Modified from the NEEP EM&V Protocols 2010)

RESIDENTIAL NATURAL GAS BOILERS AND FURNACES **Summary of Recommended EM&V Methods**		
This category is limited to residential natural gas boilers and furnaces and excludes: space heating equipment such as portable or room space heaters; electric or oil space heating equipment; and associated controls such as boiler reset controls. This category addresses stand-alone heating equipment and excludes natural gas boilers/furnaces installed through comprehensive new construction programs.		
Characteristic	**Approach**	**Additional Comments**
Program Tracking	Initial gross energy and demand savings, initial net impacts as applicable. Number of installed units, unit capacity, baseline and installed efficiency and full load heating hours.	Any additional parameters that could be useful for quality control or for evaluation design, such as sampling that are described in the EM&V plan.
Recommended M&V Method	Billing analysis (Option C) supported by telephone surveys or on-site inspections. Telephone surveys supplemented by rebate forms can confirm installation and gather data on household demographics and other operational characteristics to support the billing analysis.	Validity of billing analysis depends on whether the baseline and post-installation operation is similar, and/or appropriate corrections are made.
Alternative M&V Methods	Adding on-site inspections enhances overall confidence in household characteristics and supports collection of equipment nameplate data. Basic short-term measurements (Option A) may be added on electrical support equipment such as furnace fans and boiler pumps to refine savings estimates.	Metering methods would include time-of-use CT Loggers and spot power measurements.
	Calibrated simulation modeling (Option D) is a viable approach and is well suited for evaluating measures in a comprehensive package.	Natural gas sub-meters may be installed to isolate the heating equipment from other end uses. Collecting both electric and gas usage can be helpful in calibrating and validating building energy models.

(Source: Modified from the NEEP EM&V Protocols 2010)

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

COMMERCIAL MEASURES

Commercial measures covered here include:

- C&I Comprehensive Multi-Measure New Construction
- C&I Custom Measures
- C&I Natural Gas Boilers And Furnaces
- C&I HVAC: Prescriptive Chillers
- C&I HVAC: Unitary/Split
- C&I HVAC: Other Measures
- C&I Lighting (New Construction)
- C&I Lighting (Retrofit)
- C&I Motors
- C&I Variable Speed Drives

C&I COMPREHENSIVE MULTI-MEASURE NEW CONSTRUCTION **Summary of Recommended EM&V Methods**		
This category is limited to the installation of commercial and industrial comprehensive multi-measure new construction projects		
Characteristic	**Approach**	**Additional Comments**
Program Tracking	• Initial estimates of gross energy and demand savings, initial net impacts as applicable. • Savings by measure component; description of individual measures with, as applicable, unit quantities, sizes/capacities, baseline and installed efficiencies, and operating hours.	Any additional parameters that could be useful for quality control or for evaluation design, such as sampling that are described in the EM&V plan.
Recommended M&V Method	Calibrated simulation modeling (Option D) which is effective at capturing measure interaction. On-site data collection and review of construction documents would gather parameters, specifications, and operational characteristics to inform the model. Data collected from building Energy Management Systems (EMS) can also provide cost-effective information and should be included in EM&V plans if available.	Metering should include whole premise interval kW recorders with some end-use metering.
Alternative M&V Methods	An alternative would be to conduct on-site inspections with metering that encompasses the entire set of measures (Option B). A detailed engineering spreadsheet model can be used to capture the dynamics and interactions on an hourly basis. Less rigorous metering (Option A) could be performed but may come at the cost of reduced accuracy and validity.	Metering can be used to calibrate the model. Such metering may include whole premise interval kW recorders with some end use metering.

(Source: Modified from the NEEP EM&V Protocols 2010)

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

C&I CUSTOM MEASURES **Summary of Recommended EM&V Methods**		
This category is limited to the installation of Commercial and Industrial (C&I) custom measures in both retrofit and new construction situations. The custom category includes measures that either do not comply with or benefit from examination beyond a prescriptive calculation approach. In general, these are more complex measures that require site-specific information and detailed calculations to estimate energy and demand savings. In this context, custom measures may entail any end use or technology.		
Characteristic	**Approach**	**Additional Comments**
Program Tracking	• Initial gross estimates of energy and demand savings and initial net impacts as applicable. • Measure description with, as applicable, unit quantities, sizes/ capacities, baseline and installed efficiencies, and operating hours.	Any additional parameters that could be useful for quality control or for evaluation design, such as sampling that are described in the EM&V plan.
Recommended M&V Method	On-site inspections with partial (Option A) or complete (Option B) measurements on a census or sample of program participants. Site visits with short-term metering is the most appropriate approach for C&I Custom measures. A detailed engineering spreadsheet model can be used to capture the dynamics and interactions on an hourly basis. Data collected from Energy Management Systems (EMS) may also provide cost-effective information and should be included in EM&V plans if available.	Metering methods often include time-of-use loggers, interval kW recorders, and spot power measurements.
Alternative M&V Method	If the Custom measure involves significant HVAC equipment and/or controls, calibrated simulation modeling (Option D) offers a viable alternative for capturing measure dynamics and interaction.	Metering can be used to calibrate the model. Such metering may include whole premise interval kW recorders with some end use metering.

(Source: Modified from the NEEP EM&V Protocols 2010)

C&I NATURAL GAS BOILERS AND FURNACES **Summary of Recommended EM&V Methods**		
This category is limited to commercial natural gas boilers and furnaces.		
Characteristic	**Approach**	**Additional Comments**
Program Tracking	• Initial gross energy and demand savings and initial net impacts as applicable. • Number of installed units, unit capacity, baseline and installed efficiency, and full load heating hours.	Any additional parameters that could be useful for quality control or for evaluation design, such as sampling that are described in the EM&V plan
Recommended M&V Method	Billing analysis (Option C) supported by telephone surveys and supplemented by rebate forms and/or on-site inspections. Telephone surveys may be used to confirm installation and gather data on facility size and operating hours to support the billing analysis. Data collected from building Energy Management Systems (EMS) can also provide cost-effective information and should be included in EM&V plans if available.	Billing analysis is only valid when the pre-existing (gas bills from the pre-retrofit period) is the appropriate baseline to be used in impact analysis.
Alternative M&V Methods	Adding on-site inspections to the basic method above improves confidence in building characteristics and supports collection of equipment nameplate data. Basic short-term measurements (Option A) may be added on electrical support equipment such as furnace fans and boiler pumps to refine savings estimates.	Metering methods would include time-of-use CT Loggers and spot power measurements.
	Calibrated simulation modeling (Option D) is a viable alternative that maybe useful if Option C is inadequate or for measures that are part of a comprehensive package.	Natural gas sub-meters can be installed to isolate the heating equipment from other end uses.

(Source: Modified from the NEEP EM&V Protocols 2010)

C&I HVAC: PRESCRIPTIVE CHILLERS **Summary of Recommended EM&V Methods**		
This category is limited to air-cooled and water-cooled chiller installations in commercial and industrial facilities as a prescriptive measure. Custom chiller installations are covered under C&I Custom Measures.		
Characteristic	**Approach**	**Additional Comments**
Program Tracking	• Estimates of initial gross energy and demand savings and initial net impacts as applicable. • Number of installed units, chiller capacity, baseline and installed efficiency, and full load cooling hours.	Any additional parameters that could be useful for quality control or for evaluation design, such as sampling that are described in the EM&V plan
Recommended M&V Method	On-site inspections with partial measurements on a sample of program participants (Option A). Site visits with short-term metering offer the most cost-effective approach to prescriptive chiller projects. Data collected from building EMS may also provide cost-effective information and could be included in EM&V plans if available. Other factors that should be examined for cooling towers include the cleanliness of the cooling tower and the water temperature.	Metering methods include interval amp/kW recording or time-of-use loggers coupled with spot power measurements.
Alternative M&V Methods	An enhanced alternative is to conduct on-site inspections with metering that fully captures the entire chiller water system including supporting pumps and tower fans (Option B). Engineers can analyze hourly energy consumption for baseline and installation conditions in a dynamic spreadsheet model using Typical Meteorological Year (TMY) data.	Additional parameters of value include supply and return water temperature and water flow expressed in gallons/minute.
	Calibrated simulation modeling (Option D) is a viable alternative that is especially effective at capturing measure interaction. Simulation modeling is particularly good at temperature dependent equipment, but requires a wealth of building and operational characteristics for an accurate model.	Metering can be used to calibrate the model. Such metering may include whole premise interval kW recorders with some end use metering.

(Source: Modified from the NEEP EM&V Protocols 2010)

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

C&I HVAC: UNITARY/SPLIT **Summary of Recommended EM&V Methods**		
This category is limited to unitary HVAC installations in commercial and industrial facilities as a prescriptive measure. Unitary equipment covers split system AC, packaged systems, air-source heat pumps, and water source heat pumps. Custom unitary air conditioning applications are covered under C&I Custom Measures.		
Characteristic	**Approach**	**Additional Comments**
Program Tracking	• Initial gross energy and demand savings, initial net impacts as applicable. • Number of installed units, HVAC unit capacity, baseline and installed efficiency, and full-load cooling *and heating* hours.	Any additional parameters that could be useful for quality control or for evaluation design, such as sampling that are described in the EM&V plan.
Recommended M&V Method	On-site inspections with partial measurements on a sample of program participants (Option A). Site visits with short-term metering can offer the most cost-effective approach to prescriptive unitary/split projects. Data collected from building Energy Management Systems (EMS) can also provide cost-effective information and should be included in EM&V plans if available.	Metering methods include interval amp/kW recording or time-of-use loggers coupled with spot power, flow and temperature measurements.
Alternative M&V Methods	An enhanced alternative is to conduct on-site inspections with metering that fully surround the measurement boundary (Option B). Engineers can analyze hourly energy consumption for baseline and installation conditions in a dynamic spreadsheet model using Typical Meteorological Year (TMY) data.	Interval kW metering on whole-package units or both indoor/outdoor components of a split system.
	Calibrated simulation modeling (Option D) is a viable alternative that is effective at capturing measure interaction. Simulation modeling is particularly useful for assessing temperature dependent equipment, but requires a wealth of building and operational characteristics for an accurate model. This is a viable option for buildings with many HVAC units, zones, or solar coupling effects.	Metering would conform Option D inputs and outputs and may include the whole-premise interval kW with some space temperatures.

(Source: Modified from the NEEP EM&V Protocols 2010)

C&I HVAC: OTHER MEASURES **Summary of Recommended EM&V Methods**		
This focuses on the Other HVAC category to HVAC control measures such as thermostats, economizers, and dual enthalpy controls. This category is limited to prescriptive installations in commercial and industrial facilities. Custom HVAC applications are covered under C&I Custom Measures.		
Characteristic	**Approach**	**Additional Comments**
Program Tracking	• Initial gross energy and demand savings and initial net impacts as applicable. • Number of installed units, unit capacity and efficiency, full load cooling hours, free cooling/setback hours.	Any additional parameters that could be useful for quality control or for evaluation design, such as sampling that are described in the EM&V plan
Recommended M&V Method	On-site inspections with limited measurements on a sample of program participants (Option A). Site visits for HVAC control measures focus upon accurately inspecting and verifying operation of the controls. Data collected from building Energy Management Systems (EMS) can also provide cost-effective information and should be included in EM&V plans if available.	Metering methods may include strategically-placed time-of-use loggers to verify controls.
Alternative M&V Methods	An enhanced alternative would be an on-site inspection with metering that fully captures the impacts of the control (Option B). An hourly impact analysis would isolate the control impacts from the monitored data stream and assess across a Typical Meteorological Year (TMY) dataset.	Metering would be interval kW measurements on the affected HVAC units. Advanced metering can include enthalpy readings and damper position.
	Calibrated simulation modeling (Option D) is a viable alternative that is effective at capturing measure interaction and also control schema. Simulation modeling requires a wealth of building and operational characteristics for an accurate model. This is a viable option for buildings with many HVAC units and complex controls.	Metering would conform to Option B and include whole premise interval kW recording with some space temperatures.

(Source: Modified from the NEEP EM&V Protocols 2010)

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

C&I LIGHTING (NEW CONSTRUCTION) **Summary of Recommended EM&V Methods**		
This category encompasses C&I lighting in new construction programs.		
Characteristic	**Approach**	**Additional Comments**
Program Tracking	• Initial gross energy and demand savings and initial net impacts as applicable. • Installed quantity and wattage, corresponding baseline, fixture location, annual operating hours, in-service rate, HVAC interaction factor.	Any additional parameters that could be useful for quality control or for evaluation design, such as sampling that are described in the EM&V plan. Fixture location is critical for evaluation.
Recommended M&V Method	On-site inspections with partial measurements on a sample of program participants (Option A). Complete inspection and count of all installed lighting with spot verification of lamp/ballast type. Characterize cooling/heating zones and equipment for assessment of HVAC interactive effects. Analysis with simple engineering models. Data collected from building Energy Management Systems (EMS) can also provide cost-effective information and should be included in EM&V plans if available.	Time-of-use lighting loggers on a broad sample of fixtures, typically stratified by savings, room type, and or operating schedule.
Alternative M&V Methods	Some C&I Lighting installations warrant very high, in-building sample rates or advanced interval metering (Option B). Examples include private office spaces with high uncertainty/ diversity, hotel rooms/dormitories, and lighting systems with extensive controls. Interval kW meters are useful for recording lighting loads on circuits with many, individual occupancy sensors or dimming controls. Additional analysis with simple engineering models or 8,760 hour spreadsheets for rigorous assessment of coincident impacts could also be used.	More liberal use of lighting loggers. Alternatively, many commercial buildings isolate lighting systems in 277V power panels that can offer an excellent opportunity for interval metering on large amounts of lighting.

(Source: Modified from the NEEP EM&V Protocols 2010)

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

C&I LIGHTING (RETROFIT) **Summary of Recommended EM&V Methods**		
This category encompasses C&I lighting in retrofit programs.		
Characteristic	**Approach**	**Additional Comments**
Program Tracking	Initial gross energy and demand savings estimates and initial net impacts as applicable. Installed quantity and wattage, corresponding Baseline, fixture Location, annual operating hours, in-service rate, HVAC interaction factor.	Any additional parameters useful for quality control such as sampling. Fixture location is critical for evaluation.
Recommended M&V Method	On-site inspections with partial measurements on a sample of program participants (Option A). Complete inspection and count of all installed lighting with spot verification of lamp/ballast type. Characterize cooling/heating zones and equipment for assessment of HVAC interactive effects. Analysis with simple engineering models. Data collected from building EMS may also provide cost-effective information and should be included in EM&V plans if available.	Time-of-use lighting loggers on a broad sample of fixtures, typically stratified by savings, room type, and or operating schedule.
Alternative M&V Methods	Some C&I lighting installations warrant very high, in-building sample rates or advanced interval metering (Option B). Examples include private office spaces with high uncertainty/diversity, hotel rooms/dormitories, and lighting systems with extensive controls. Interval kW meters have proven useful for recording load on lighting circuits with many, individual occupancy sensors or dimming controls. Analysis with simple engineering models or 8,760 hour spreadsheets for rigorous assessment of coincident impacts could also be used.	More liberal use of lighting loggers. Alternatively, many commercial buildings isolate lighting systems in 277V power panels that offer a prime opportunity for interval metering on large amounts of lighting.

(Source: Modified from the NEEP EM&V Protocols 2010)

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

C&I MOTORS **Summary of Recommended EM&V Methods**		
This category is limited to the installation of premium efficient motors in C&I facilities as a prescriptive measure.		
Characteristic	**Approach**	**Additional Comments**
Program Tracking	• Initial gross energy and demand savings and initial net impacts as applicable. • Number of installed units, motor horsepower, ends Use and application (e.g. HVAC supply fan), Location, baseline and installed efficiency, Loading factor, and annual operating hours.	Any additional parameters useful for quality control and also for evaluation design such as sampling. Motor location is critical for evaluation.
Recommended M&V Method	On-site inspections with partial measurements on a sample of program participants (Option A). Basic site visits with time-of-use metering offers the most defensible and cost-effective approach to constant-speed, prescriptive motors. Data collected from building EMS may also provide cost-effective information and should be included in EM&V plans if available.	Metering methods include time-of-use CT or "magnetic field" loggers and spot power measurements.
Alternative M&V Methods	An enhanced alternative is to conduct on-site inspections with interval kW metering that track the electrical performance of the motor throughout its load range (Option B). This added rigor captures part-load efficiency effects that tend to be neglected in a Time-of Use (TOU) metered approach with SEMs.	Metering would be interval kW measurements for a reasonable duration to span a variety of motor loading situations.

(Source: Modified from the NEEP EM&V Protocols 2010)

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

C&I VARIABLE SPEED DRIVES **Summary of Recommended EM&V Methods**		
This category is limited to Variable Speed Drives (VSD) installations in C&I facilities as a prescriptive measure. Custom VSD applications are covered under C&I Custom Measures.		
Characteristic	**Approach**	**Additional Comments**
Program Tracking	• Initial gross energy and demand savings and initial net impacts as applicable. • Number of installed units, motor horsepower, end use and application (e.g. HVAC supply fan), Location, savings factors, and annual operating hours.	Any additional parameters useful for quality control and also for evaluation design such as sampling. VSD location is critical for evaluation.
Recommended M&V Method	On-site inspections with interval kW metering that tracks the electrical performance of the motor/VSD combination throughout its load range (Option B). Lesser rigor would not capture the variability intrinsic to a VSD application.	Metering would be interval kW measurements for a reasonable duration to span a variety of loading situations.
Alternative M&V Methods	Calibrated simulation modeling (Option D) is an alternative that is effective at measure interaction and control schema. Simulation modeling requires a wealth of building and operational characteristics for an accurate model. This is a viable option for facilities with many VSDs on HVAC systems units.	Metering can be used to calibrate and validate the model. Such metering would mirror Option B perhaps with whole premise interval kW recording and some space temperatures.

(Source: Modified from the NEEP EM&V Protocols 2010)

PROTOCOL E: PROTOCOLS FOR VERIFICATION AND ONGOING MODIFICATION OF DEEMED SAVINGS VALUES

Protocol Scope: These protocols include the recommended timing for updating deemed savings values, especially for those technologies in which significant energy efficiency improvements occur on a periodic basis, such as lighting. These protocols also provide sources to consider when reviewing and modifying deemed savings, based on the findings from the literature review.

Customer Classes: All except self-directing customers

PROTOCOL E1: REVISING AND UPDATING DEEMED SAVINGS VALUES

1. Each deemed measure or measure set in the Arkansas Technical Reference Manual (TRM) is subject to a review to establish a "Sunset Date" at a minimum of every three (3) years or sooner if sooner conditions warrant. High Impact Measures will be reviewed annually. A High Impact Measure (HIM) is an energy efficiency measure that accounts for at least 5% of total portfolio gross kilowatt hour, kilowatt, and/or therm savings in one or more of the utility's energy efficiency programs.

2. Upon reviewing a Deemed Measure, the parties participating in the parties working collaboratively process may recommend that the Commission:
    a. Extend the "Sunset Date" for the measure with its cost and savings unchanged;
    b. Adopt revised cost and savings assumptions for the measure;
    c. Re-instate the Deemed Measure, contingent on the outcome of future evaluations, M&V, engineering work, and/or market research;
    d. Sunset the Deemed Measure and recommend that
        - A Simplified M&V Protocol will be developed for it;
        - It will be treated as a Custom M&V Measure; or
        - It will be eliminated.

PROTOCOL E2: IMPLEMENTATION OF CODE CHANGES

Codes and standards that affect equipment and systems sold and installed in Arkansas may be periodically updated. These codes and standards may include, but are not limited to, those listed below:

1. The International Energy Conservation Code, IECC
2. ASHRAE 90.1 - Energy Standard for Buildings Except Low-Rise Residential Buildings
3. Arkansas Energy Code
4. The National Appliance Energy Conservation Act (NAECA)
5. EISA – the Energy Independence and Security Act of 2007

The purpose of these code and standards updates is to increase energy efficiency by codifying minimum equipment performance or baselines. This is in contrast to DSM programs, which are intended to motivate participants to install systems and equipment that exceed prevailing codes or standards.

These code and standard changes affect and change the baseline efficiency or performance used in calculating savings for; (1) replace on burnout and (2) new construction projects as well as (3) 'outyear' baselines on dual-baseline early retirement projects.

It is recognized that there is a lag between the time when a code or standard comes into force and when the industry has made a substantial transition to the new code or standard. In recognition of this lag time, the TRM provides a lag time to allow for industry to adjust. Specifically, the TRM adoption date for a code or standard update is the end of the current program year if the effective date of the code or standard update is before July 1. For code or standard effective dates on or after July 1, the enforcement date is the end of the following program year.

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

PROTOCOL F: PROTOCOLS FOR THE DETERMINATION OF NET PROGRAM IMPACTS

Protocol Scope: These protocols, commonly used to determine net-to-gross ratio ("NTGR"), isolate free ridership and spillover rates. This protocol is designed to clarify the steps necessary to complete a "true-up" of program savings estimates ex post to determine the Lost Contribution to Fixed Costs (LCFC).

Customer Classes: All except self-directing customers

There are five approaches for determining NTGR:

- **Self-Reporting Surveys:** From participants and non-participants without independent verification.
- **Enhanced Self-Reporting Surveys:** 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:** Statistical models are used to compare participant and non-participant energy and demand patterns. These models often include survey inputs and other non-program-related factors such as weather and energy costs (rates).
- **Deemed Net-to-Gross Ratios:** An NTGR is estimated using information available from evaluation of similar programs.
- **Stipulation of Net-to-Gross Ratios:** The stipulation of a net-to-gross ratio may be appropriate when the expense of the NTGR analysis and the uncertainty of the results are considered significant barriers (NAPEE 2007).

Recommended Net-to-Gross Analysis

Net-to-Gross (NTG) analysis is an important component of program evaluation because it helps to quantify estimated savings attributable to a program. NTG estimation involves triangulating data from multiple sources and therefore is incorporated into both the process and impact evaluation tasks. Relevant data sources include: surveys or in-depth interviews with customers, trade allies, and other key program stakeholders; data collection during on-site field inspections; billing records; and demand elasticity modeling using participating retailer sales data. In addition, NTG estimates may be further validated through a comparison of results from similar energy efficiency programs operating in other jurisdictions or examining a range of market data sources such as surveys, conference proceedings, and market assessments.

Estimating the impacts that are attributable to a program poses many challenges. First, the participants may not be able to accurately answer the necessary hypothetical question: *"What measures would you have installed anyway if you had not participated in the program?"* However, attribution is estimated not only for energy efficiency programs but also for other public policy initiatives/investments. While this is difficult and subject to some judgments and assumptions, NTG analyses can be performed on energy efficiency programs in such a way that reasonable information can be provided to policy makers that will assist them in making good decisions about these programs. While this is not an exact science, the attribution methods proposed herein are meant to achieve that objective (i.e., provide information in context that will inform policymakers and assist in assessing historic and future investments in energy efficiency).

Methodology

This section presents general definitions and methods that will be employed as part of a sound NTG analysis. The discussion is purposefully kept at a relatively high-level; additional details regarding the question sets and methods used to conduct the NTG analysis will be provided by the EM&V contractor. The NTG calculation will be applied retrospectively to the gross savings achieved during the program year being evaluated.

Derivation and Definition of Attribution

The methodology for assessing the energy savings attributable to a program is based on a NTGR that has two main components: free ridership and spillover.

Free ridership refers to program participants who received an incentive but would have installed the same efficiency measure on their own had the program not been offered. This includes partial free riders, defined as customers who, at some point, would have installed the measure anyway, but the program persuaded them to install it *sooner* or customers who would have installed the measure anyway but the program persuaded them to install more efficient equipment and/or more equipment. For the purposes of EM&V activities, participants who would have installed the equipment within one year will be considered full free riders; participants who would have installed the equipment later than one year will not be considered to be free riders (thus no partial free riders will be allowed).

Free ridership is the share of gross program savings that is generally the savings accounted for in program records and then adjusted for the naturally occurring adoption; the free ridership rate is based on actions participants "would have taken anyway" (i.e., actions that were not induced by the program). Each energy efficiency program covers a range of energy efficiency measures and is designed to move the overall market for energy efficiency forward. However, it is likely that some participants would have wanted to install some high efficiency measures (possibly a subset of those installed under the program) even if they had not participated in the program or been influenced by the program in any way.

Spillover refers to energy savings that are due to the influence of a program but are not counted in program records. For example, a customer installs a set of efficiency measures in one of his/her buildings. These measures were promoted (and incented) under a DSM program. The customer then decides to install the same measures at another site, where there is no program incentive. In this case, the program had an influence on the market beyond the energy savings in this customer's first building. Spillover can be broken out in three categories:

- **Participant Internal Spillover** represents energy savings from additional measures implemented by participants at participating sites not included in the program but directly attributable to the influence of the program.

- **Participant External Spillover** represents energy savings from measures taken by participants at non-participating sites not included in the program but directly attributable to the influence of the program.

- **Non-Participant Spillover** represents energy savings from measures that were taken by non-participating customers but are directly attributable to the influence of the program.

Spillover adds to a program's measured savings by incorporating indirect (i.e., not incented) savings and effects that the program has had on the market above and beyond the directly incented or directly induced program measures.

Total spillover is a combination of several factors that may influence non-reported actions to be taken at the project site itself (inside spillover) or at other sites by the participating customer (outside spillover). Each type of spillover is meant to capture a different aspect of the energy savings caused by the program, but not included in program records. Because a primary goal of most DSM programs is to transform markets through a variety of strategies – including education, promotion, and increasing awareness of the benefits of energy efficiency – one would expect spillover to occur to some extent in the market.

The overall NTGR is meant to account for both the net savings at participating projects and spillover savings that result from the program (but are not included in program records). When the gross program savings multiplies the NTG ratio, the result is an estimate of energy savings that are attributable to the program (i.e., savings that would not have occurred without the program). The basic equation is:

$$NTG = 1 - Free\ ridership + Spillover$$

The underlying concept inherent in the application of the NTG formula is that only savings caused by the program should be included in the final net program savings estimate, but this estimate should include all savings caused by the program (i.e., the net program savings should account for free ridership and include spillover).

Estimating Free Ridership: Survey Techniques

Data to assess free ridership should be gathered through a series of survey questions asked of end-use customers and trade allies who participated in the program. Free ridership can be evaluated by asking direct questions, aimed at obtaining respondent estimates of the appropriate free ridership rate that should be applied to them, and by supporting, or influencing questions used to verify whether the direct responses are consistent with participants' views of the program's influence.

The direct free ridership questions ask respondents to estimate the share of measures that would have been incorporated at high efficiency if not for the technical and financial assistance of the program. The questions also ask respondents to estimate the likelihood that they would have incorporated measures "of the same high level of efficiency" if not for the technical and financial assistance of the program. This flexibility in how respondents conceptualize and convey their views on free ridership will allow respondents to provide their most informed response, thus improving the accuracy of the free-ridership estimates.

The "program influence" questions clarify the role that program interventions (e.g., financial incentives and technical assistance) played in decision-making and provide supporting information on free ridership. Responses to these questions are analyzed for each respondent and used to identify whether the direct responses on free ridership are consistent with how each respondent rated the "influence" of the program.

These results will then be compared to free ridership estimates based on on-site inspections/audits, and/or estimates derived from similar surveys completed in other jurisdictions.

Estimating Spillover: Survey Techniques

The basic method for assessing participant (inside and outside) spillover employs a three-step approach to determine the following:

1. **Whether spillover exists at all**. These are yes/no questions that ask, for example, whether the respondent incorporated energy efficiency measures or designs that were not recorded in program records. Questions relate to extra measures installed at the project site (inside spillover) and to measures installed in non-program projects (outside spillover).

2. **The extent of the spillover.** These questions request information about the number or share of projects/jobs/facilities into which additional measures or technologies are installed (these questions are not asked for inside spillover because the value is simply the one project on which the interviewee focuses).

3. **The amount of savings per spillover project.** These questions ask respondents to estimate the energy savings associated with the non-recorded measures relative to the savings from the participating project itself.

The outcome of these inquiries is an estimate of the share of those non-recorded savings that can be attributed to the influence of the program.

Timing of Data Collection for Free Ridership vs. Spillover

The evaluation team should, where possible, use a staggered data collection approach to collect information in support of the NTG analysis. The rationale for this approach is that free ridership and spillover data are best collected at different points in time. Free ridership data are considered to be most accurate when collected as closely as possible to the point in time when the participation decision is made. Doing so helps to ensure accurate participant recall of motivating factors and relative program influence while also producing other benefits, including near-term feedback for program staff regarding program influence effects. Conversely, spillover data are considered most accurate when collected sometime after the participating project has been completed. Allowing a reasonable amount of time to pass before asking participants about spillover effects ensures that participants have sufficient time to: a) install the incented equipment, b) experience its operating parameters and costs, and c) then decide whether or not to install additional energy efficiency measures at the project site or some other location independent of any program support or financial incentive (Johnson et al., 2010).

Hierarchical Approaches for Determining When to Update NTG Values

One of the primary goals of this protocol is to provide a common framework with respect to the updating of NTG. It is critical that all utilities and their evaluators adopt this protocol to ensure a more thorough and consistent approach to net savings estimation and application, plus to direct evaluation resources to the areas of highest uncertainty. Note that the intent of this protocol is *not* to dictate the specifics (e.g., survey batteries or algorithms, market based approaches, etc.) with which NTG is estimated, but rather to help determine when an updated NTG estimate is needed.

A several step decision tree should be used to help steer the timing for updating attribution analysis, which in the future should help evaluators determine when to collect current primary data, and when prior research or Arkansas specific secondary data might be reliable, with the final option being to rely on literature reviews. The framework is straightforward, whereby the updating of net savings follows the hierarchical approach (presented visually in Figure 5):

1. **Has NTG research been conducted on the same program in a prior year?** The first step to determining whether primary NTG research should be conducted in a given program year is to assess whether primary data – collected for the same program – are available from a prior year.[4] If prior data are available, the evaluation contractor should

[4] Note that the prior data do not necessarily need to be from the prior year, nor do they need to necessarily be from a single year (e.g., two years of data may also be used if both years are determined to meet these criteria and it is believed the additional data would provide more robust results).

determine whether the prior values are applicable in the current year. There are at least two overarching components of this decision, both of which could have significant impacts on the NTG ratio.[5]

a. First, determine if the current program is similar to the program in which the primary data was collected: Is the mix of measures the same? Is the contribution to savings for each measure similar? Are the incentive levels comparable? Is (are) the delivery method(s) similar?

b. Second, determine if the market conditions are similar to the time period in which the prior data were collected: Has there been a substantial change in incremental cost for the efficient measures? Has there been a substantial change in the supply or availability of the efficient measures? Has there been a substantial change in the market share of efficient measures (i.e., the ratio of efficient measures sold to total comparable standard and efficiency measures)? Are the local or federal codes and standards the same as when the prior NTG values were estimated?

If the program and market conditions are comparable to the time period(s) in which the prior primary NTG research was conducted, these prior values can be considered applicable to the current program year.

2. **If prior year primary data are not available or are determined not to be applicable due to changes to either program or market conditions.** The evaluator should then determine whether or not the estimated savings from the program support primary research. In general, programs that represent at least 5-10 percent of the portfolio estimated savings in any given year should use NTG ratios that are estimated via primary data research for that specific program.[6]

3. **If prior year data for the program are not available or applicable, and the program savings does not support primary data collection.** The evaluation should then consider if NTG values derived from Arkansas-based comparable programs are available. A comparable program is defined as one that is similar in terms of program maturity, incentive levels, delivery mechanism, and measure types. Ideally, NTG values derived in the same program year would be used, but values from prior years may also be used if the comparability conditions are met.

4. For existing and new programs that do not meet any of the above specifications, then a literature review may be undertaken to locate a similar program (or programs) that has (or have) an established NTG value(s). This approach requires that the research be well documented, ***and the selected NTG value be reviewed and agreed to as reasonable by the IEM***. A program may be identified as similar if it meets the following conditions:

[5] Note the questions presented here are examples of questions to ask to assess comparability, but are not meant to be exhaustive of the types of questions that should be addressed.

[6] The IEM understands that smaller portfolios have limited budgets to support any primary research, these thresholds are estimates and may vary based on the overall portfolio budget. As discussed, final decisions regarding the use of primary vs. secondary data for NTG research should be made in consultation with the IEM as part of the evaluation planning process.

i. **Program similarity**: maturity, incentive levels, delivery mechanism, and measure types are similar
ii. **Market similarity:** demographic, household, and business characteristics are similar (or as similar as possible) to Arkansas

The IEM believes this hierarchical approach maximizes the use of valuable evaluation resources towards programs that could most benefit from primary research, and thus avoids unnecessarily repeating NTG research every year for the same programs. However, to prevent NTG values from being repeated too many years and becoming potentially "stale", NTG values for programs that meet the contribution to savings threshold (#2 above) should be updated at least via primary research at least once during every three-year program cycle.

The IEM also understands that these decisions are open to some amount of interpretation and subjectivity (e.g., determining exactly what constitutes a substantial change in incremental cost, availability, or market share), so the steps along this decision tree should be clearly presented and discussed as part of the annual evaluation plans, and thus decision can be made in consultation with the IEM. The EM&V planning process has typically occurred in the summer months for the current program year. The IEM, however, is also available to discuss and agree upon final evaluation NTG approaches earlier in each program year.

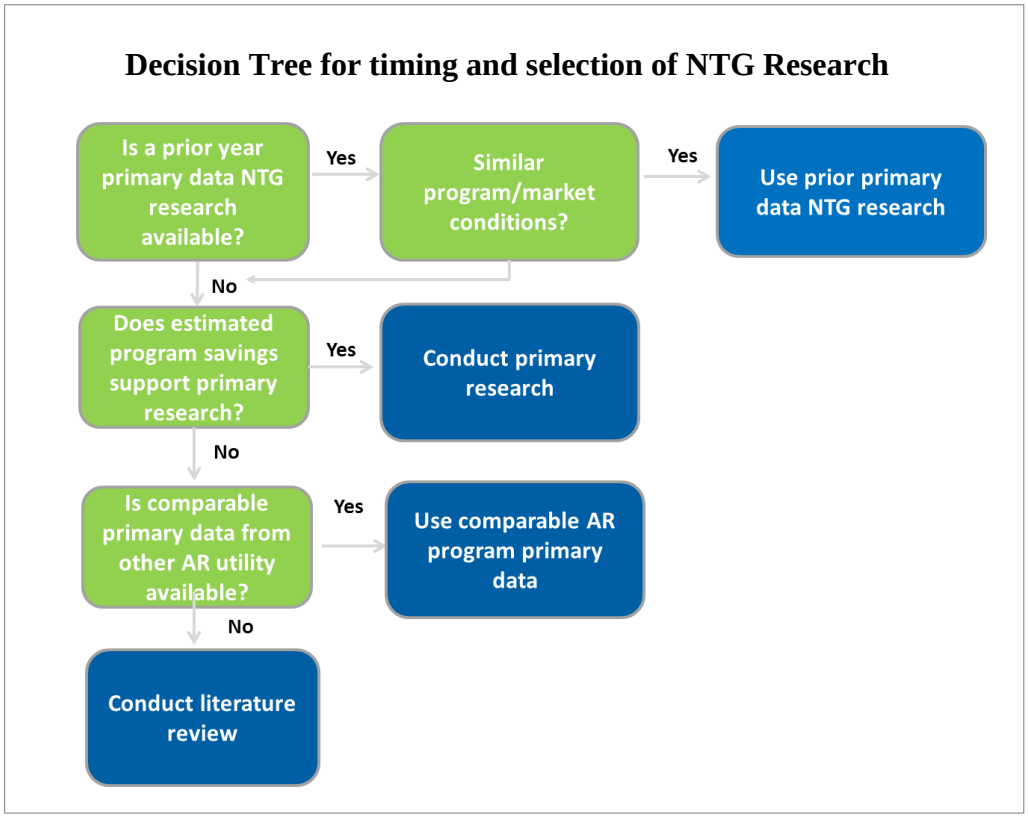


Figure 5: Decision Tree for Timing and Selection of NTG Research

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

Application of Trade Ally Input

As noted in the PY2013 IEM Report, the PY2013 evaluations included a number of NTG estimates that – according to the program evaluation plan – were going to leverage trade ally input, yet did not include trade ally responses or values (IEM PY2013 Annual Report, pg. xiii). In addition, a number of programs did utilize trade ally input, yet the research provided limited evidence that the responses were representative of the programs for which the values were utilized.

It is imperative that the evaluations adhere to the approved work plans, and where they diverge note the reasons for doing so and provide strong argument as to why they were not followed (e.g., inadequate sample, poor response rates, etc.). This is particularly true for NTG methods, which can provide widely different estimates depending on the method and the respondent type.

Evaluators should include specific details regarding their planned integration of trade ally responses with customer survey responses for overall program attribution within their work plans. Evaluations using trade ally responses should be collected for programs where the trade allies play a key role in the installation decision[7], and the work plans should present a discussion of the representation from the trade ally respondents. This protocol is not requiring evaluators to follow a specific algorithm to integrate trade ally responses, but the viability and consistency of approaches will be addressed by the IEM during the work plan review.

[7] The importance of the trade ally on the decision to install an energy efficient measure can be assessed through participant surveys.

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

PROTOCOL G: PROVISIONS FOR LARGE CUSTOMERS

Protocol Scope: This protocol provides more detailed information regarding projects installed by industrial and commercial customers, and the challenges associated with establishing a streamlined process for measurement and verification. These protocols rely on the established "best practices" for ensuring that large custom-project directors will maintain the necessary building records and measure data to provide for robust EM&V activities and accurate measurement of energy savings. Often these projects involve highly technical, on-site engineering analysis and verification, which should be performed efficiently and cost-effectively, with as little impact on the customer site as possible. It is important to note that these types of EM&V protocols address custom measures for which deemed savings values do not exist and for prescriptive/direct-install measures that are being used in a non-traditional manner.

For specific guidance on the Opt-Out/Self-Direct Option, please consult Section 11 of Conservation & Energy Efficiency (C&EE) Rules filed in Docket 10-101-R.

Customer Classes: Large C&I customers except self-directing customers

The objectives of measure installation verification are to confirm that the:

- Measures were actually installed,
- Installation meets reasonable quality standards, and
- Measures are operating correctly, have the potential to generate the predicted savings.

The M&V should also verify and quantify actual savings at the site.

Installation verification should be conducted at all sites claiming energy or peak demand impacts where M&V is conducted. Installation verification activities may also be specified by the process or market effects protocols. Data collected from the building's EMS may provide cost-effective information and should be included in EM&V plans if available.

M&V projects conducted under this protocol shall adhere to the International Performance Measurement and Verification Protocol (IPMVP).

Development of Site-Specific M&V Plan: This protocol requires submittal of an M&V plan for each field measurement project that documents the project procedures and rationale in such a way that the results can be audited for accuracy and repeatability. Within the guidelines established by the IPMVP and these protocols, there is considerable latitude for the practitioner in developing a site-specific M&V plan and implementing the plan in the field. The M&V contractor shall evaluate the uncertainty in the desired data product, and develop a site-specific M&V plan that manages the uncertainty in the most cost-effective manner.

Initial estimates of engineering parameter uncertainties should be used to provide an estimate of the overall uncertainty in the savings calculations. Assumptions used to create initial estimates of parameter uncertainty values should be documented. The contribution of specific engineering parameters to the overall uncertainty in the savings calculations should be identified and used to guide the development of the M&V plan. The components of the M&V plan should:

1. *Identify Goals and Objectives.* The goals and objectives of the M&V project should be stated explicitly in the M&V plan.
2. *Specify Site Characteristics.* Site characteristics should be documented in the plan to help future users of the data understand the context of the monitored data. Depending on the nature of the measure, the site characteristics description should include:
   a. General building configuration and envelope characteristics, such as building floor area, conditioned floor area, number of building floors, opaque wall area and U-value; window area, U-value and solar heat gain coefficient.
   b. Building occupant information, such as number of occupants, occupancy schedule, building activities.
   c. Internal loads, such as lighting power density, appliances, plug and process loads.
   d. Type and quantity and nominal efficiency of heating and cooling systems.
   e. Important HVAC system control set points.
   f. Changes in building occupancy or operation during the monitoring period that may affect results.
   g. Description of the energy conservation measures at the site and their respective projected savings.
3. *Specify Data Products and Project Output.* The planned output and results of the M&V activity should be specified. These data products should be referenced to the goals and objectives of the project, and include a specification of the data formats and engineering units.
4. *Specify an M&V Option.* The M&V option chosen for the project should be specified according to the IPMVP, consistent with the M&V protocol.
5. *Specify Data Analysis Procedures and Algorithms*. Engineering equations and stipulated values, as applicable, shall be identified and referenced within the M&V plan. Documentation supporting baseline assumptions shall be provided.
6. *Specify Field Monitoring Data Points.* The actual field measurements planned should be specified, including the sensor type, location, and engineering units.
7. *Estimate Data Product Accuracy.* All measurement systems have error, expressed in terms of the accuracy of the sensor and the recording device. The combined errors should be estimated using a propagation of error analysis, and the final data product should be accuracy described.
8. *Specify Verification and Quality Assurance Procedures.* Data analysis procedures to identify invalid data and treatment of missing data and/or outliers must be provided.
9. *Specify Recording and Data Exchange Formats*. Data formats compliant with the data reporting protocol should be described.

(Modified and Expanded from CA Evaluators' Protocols 2006).

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

Additional Guidance for the IPMVP Protocols for Custom Projects
1. **Measure existence** should be verified through on-site inspections of facilities; measure make and model number data shall be collected and compared to participant program records as applicable. Sampling may be employed at large facilities with numerous measures installed. As-built construction documents may be used to verify measures such as wall insulation where access is difficult or impossible. Spot measurements may be used to supplement visual inspections, such as solar transmission measurements and low-e coating detection instruments to verify the optical properties of windows and glazing systems.
2. Quality of Installations: Measure installation inspections shall note the quality of measure installation, including the level of workmanship employed by the installing contractor toward the measure installation and repairs to existing infrastructure affected by measure installation, and physical appearance and attractiveness of the measure in its installed condition. Installation quality guidelines developed by the program implementer shall be used to assess installation quality. If such guidelines are not available, then the guidelines shall be developed by the M&V contractor and approved by the Commission prior to conducting any verification activities. Installation quality shall be determined from the perspective of the customer.

(Source: Modified from CA Evaluators' Protocols 2006)

PROTOCOL H: TECHNICAL REFERENCE MANUAL (TRM)

Protocol Scope: To provide a clear and effective method for updating the Arkansas TRM

The Arkansas TRM is designed to be a dynamic document which will benefit from periodic updates developed by the PWC through an objective and thoughtful process. Defining a process that coordinates with the needs of users, evaluators, and the APSC is critical. It is critical to maintain a current TRM and consider any necessary updates to the document at least annually, at the same time recognizing the need for the users of the TRM and others to have some degree of certainty as to the TRM values upon which they can rely. Accordingly, this protocol describes the process for updating the TRM and coordinating this process with other critical activities. The annual update process set forth herein is the preferred course of action for updating the TRM.

TRM Update Process

The PWC should work cooperatively to identify any necessary revisions and to present any revisions to the Commission by August 31 each year. Examples of events that may precipitate the need to consider changes to the TRM, may include but are not limited to:

- **New measure additions.** As new technologies become cost-effective, they will need to be characterized and evaluated for addition to the manual. In addition, new program delivery design may result in the need for new measure characterization.
- **Existing measure updates.** Updates may be required for a number of reasons. Examples include: the federal standard for efficiency of a measure has changed; the qualification criteria are altered; the measure cost falls; or a new evaluation provides a better value of an assumption for a variable. In addition, as programs mature, characterizations need to be updated, where changes in the market support changes in calculation assumptions. In such cases, these changes must be identified and appropriate changes to the TRM evaluated.
- **Retiring existing measures.** When the economics of a measure become such that it is no longer cost-effective, or the free rider rate is so high that it is not worth supporting the measure, or if the market has changed, then the measure should be evaluated for retirement.
- **High Impact Measure (HIM) reprioritization.** The prioritization of measures in terms of the HIM Tier classification is subject to change over time depending on the relative magnitude of reported energy savings resulting from actual program measure installations.

The flowchart in Figure 6 outlines the steps for regular TRM updates. The PWC will work cooperatively to identify any necessary revisions to the TRM. To ensure there is a clear differentiation between policy and technical matters, the PWC should establish regular meetings devoted to policy issues and resolving technical issues in which the PWC can discuss these matters and determine the any necessary revisions to the TRM for recommendations to the Commission for approval.

The process outlined in Figure 6 requires a number of different roles to ensure effectiveness, sufficient review, and independence. The following is a list of key roles and responsibilities for this process. The list of roles and responsibilities is not comprehensive.

- **Arkansas Public Service Commission (Commission)**
  - Approves or denies any changes to the TRM, as well as this TRM process.

- **Independent Evaluation Monitor (IEM)**
  - Assures compliance with national Evaluation, Measurement, and Verification ("EM&V") "best practices," and Commission-approved protocols and the Arkansas TRM.
  - Manages timely updates and/or expansion of deemed savings and the TRM are pursued.
  - Oversees and coordinates the activities of the TRM Technical Manager.
  - Gives feedback on draft measure characterizations from other parties
  - Coordinates with Staff on recommendation for TRM revision to the Commission.
  - Manages and updates TRM manuals (after Commission approval of changes).
  - Ensures proper use of TRM in annual savings verification process.

- **Program Administrators / Utilities / Program Implementers**
  - Identify need for new or revised measure characterization (usually due to program changes or program/market feedback).
  - Communicates need for new or revised measure to IEM.
  - Give feedback on draft measure characterizations from other parties.
  - Participate in formal discussion and dispute resolution.

- **TRM Technical Manager**
  - Identifies need for revised measure characterization (usually based on knowledge of local or other relevant evaluation studies).
  - Reviews, researches and/or develops draft measure characterizations identified either by itself, EM&V Contractor, Utility, IEM or other party.
  - Incorporates revisions to draft and final TRM documents.

- **EM&V Contractors**
  - Identify need for revised measure characterization (usually based on local evaluation studies it has conducted or managed).
  - Research and/or prepare draft measure characterization for PWC consideration.
  - Provide input/feedback on draft measure characterizations developed by other parties.
  - Perform program evaluation - includes statewide market assessment and baseline studies, savings impact studies (to measure the change in energy and /or demand use attributed to energy efficiency), and other energy efficiency program evaluation activities.
  - Verify annual energy and capacity savings claims of each program and portfolio.

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

- **Staff**
  - Works with the PWC and IEM to identify any necessary changes to TRM.
  - Annually, by August 31, submits recommended revisions to the TRM to the Commission for its approval.
  - Provides supporting testimony for any recommended revisions.

- **Other Parties to the Docket/ Interveners**
  - Identify need for new or revised measure characterization (usually based on knowledge of local or other relevant evaluation studies).
  - Give feedback on draft measure characterizations from other parties.
  - Provide input and assist in identifying necessary revisions to the TRM.
  - Provide testimony as needed addressing recommended revisions to the TRM.

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

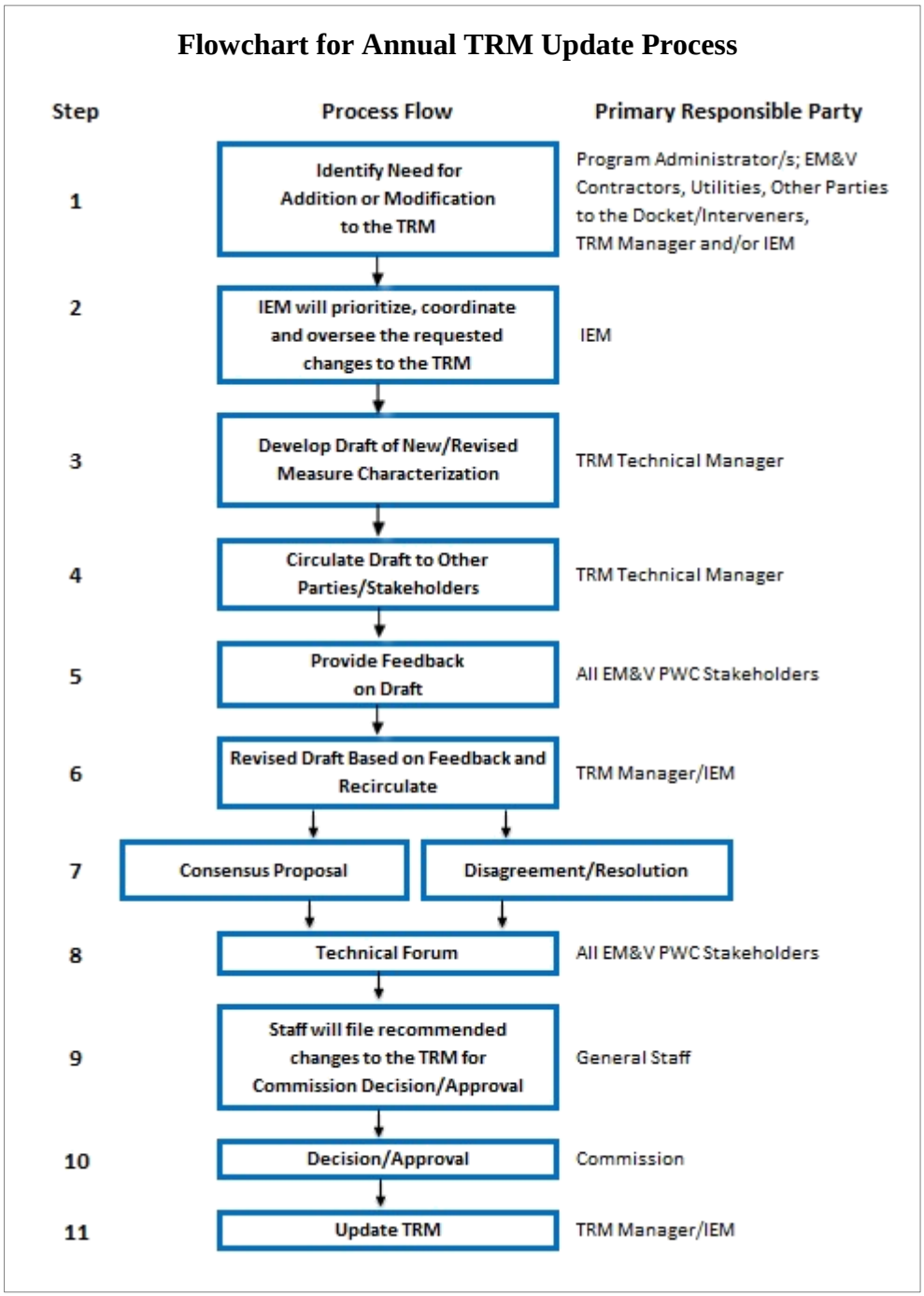


Figure 6: Flowchart for Annual TRM Update Process

This process includes several potential stages of discussion and feedback on draft modifications to the TRM. The IEM will convene a Technical Forum for the PWC and other key stakeholders. This forum will identify the changes made to the annual TRM update, and highlight the findings in the Annual TRM update. It will also provide an opportunity for the

Parties to ask questions and provide more detailed information regarding the scope of these changes.
Items included in the Technical Forum include:

- Present what changes were made to the TRM, and the reasons for the changes including a detailed discussion of the assumptions made and the basis for these findings;
- Review the timing for incorporating the changes;
- Discuss the implications of the changes on current and future programs;
- Identify other potential energy efficient technologies that should be considered for future TRM updates, based on our experiences in other jurisdictions; and
- Attempted resolution of any disagreements.

The Technical Forum may take the form of a one-day workshop, and will include presentations from the EM&V contractors and the PWC in addition to the appropriate members of the IEM team.

Table 7 provides a recommended timeline for a coordinated process in line with the Commission deadlines.

Table 7: Annual Verification and TRM Update Timeline

<table>
<tr><th></th><th>Jan</th><th>Feb</th><th>Mar</th><th>April</th><th>May</th><th>June</th><th>July</th><th>Aug</th><th>Sept</th><th>Oct</th><th>Nov</th><th>Dec</th></tr>
<tr><td>Utility/Third Party Administrators</td><td colspan="3" rowspan="2">Prepare Utility Energy Efficiency Annual Report</td><td colspan="2" rowspan="2">Draft new/updated measure characterization and submit to IEM/TRM for Review and Comment</td><td colspan="2" rowspan="6">Participate in discussions of TRM Update; Review recommended changes Review drafts prepared by the TRM Manager (i.e., Frontier) Participate in Technical Forum for TRM Udpates;</td><td rowspan="6">Review Final TRM</td><td colspan="4" rowspan="3">Savings Verification/ EM&V Activities; Identify potential changs for consideration in future TRM Updates</td></tr>
<tr><td>EM&V Contractors</td></tr>
<tr><td rowspan="4">IEM/TRM Manager</td><td></td><td></td><td colspan="3" rowspan="4">Review EM&V Reports and TRM Update Suggestions; Priortize Recommendations for Consideration by the PWC</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="2">General Staff/Other Parties to the Docket</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Commission</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="4">Approve TRM Changes</td><td></td></tr>
<tr><td></td><td>Jan</td><td>Feb</td><td>Mar</td><td>April</td><td>May</td><td>June</td><td>July</td><td>Aug</td><td>Sept</td><td>Oct</td><td>Nov</td><td>Dec</td></tr>
</table>

Detailed Process for Updating TRM

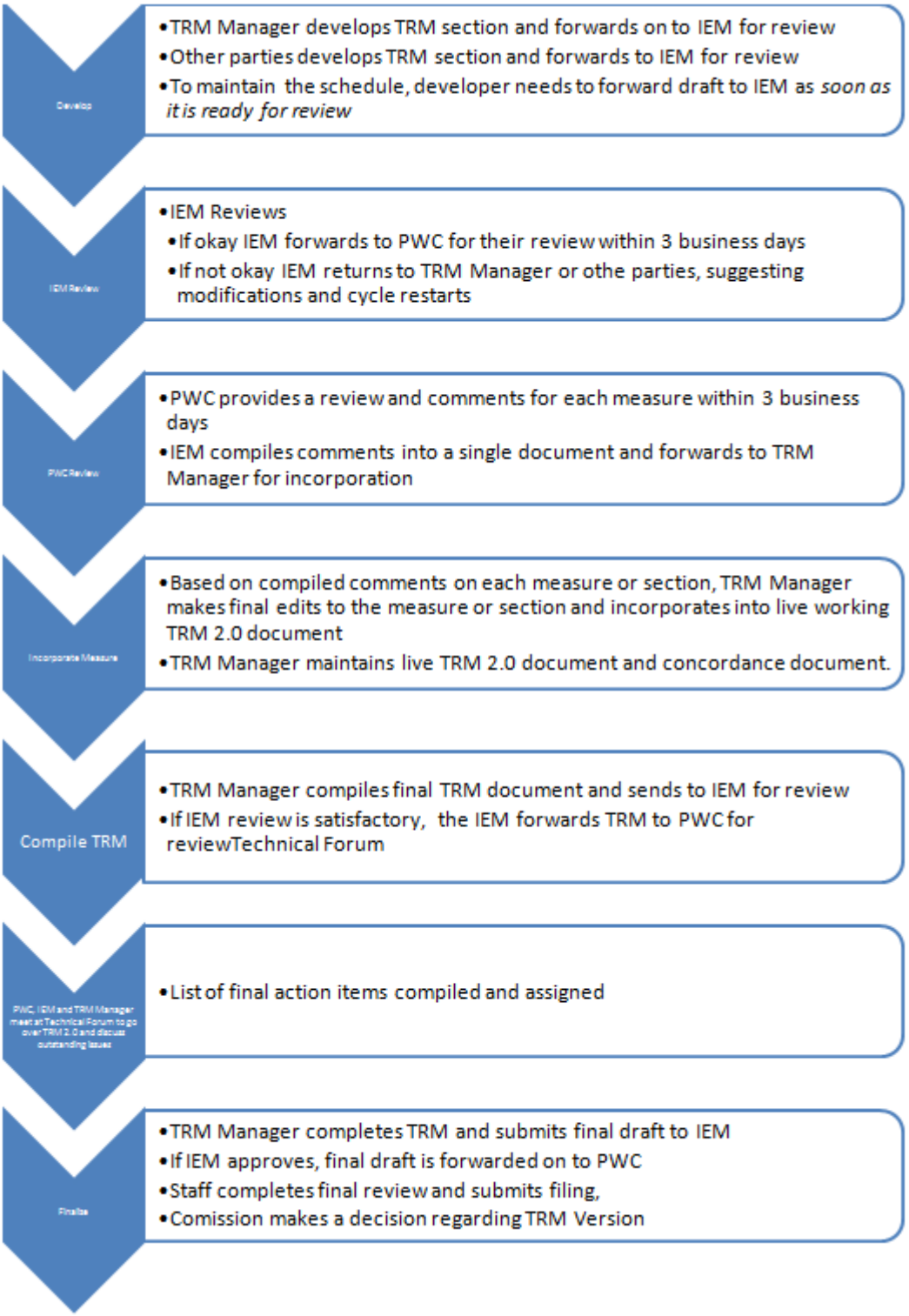


Figure 7: Detailed Process for Updating TRM

PROTOCOL I: ROLE AND RESPONSIBILITIES OF THE INDEPENDENT EM&V MONITOR

The Commission's Order, dated December 10, 2010, establishing the Parties Working Collaboratively (PWC) to develop an EM&V protocol required that Staff file with the Commission, on or before June 1, 2011, suggested EM&V rule changes requiring the implementation of EM&V in accordance with National Action Plan for Energy Efficiency (NAPEE) "best practices." Commission-ordered Task (2)(i) on page 16 calls for the "establishment of an ongoing, annual process for a single, Independent EM&V Monitor (IEM) jointly funded by EE utilities, to report to the Commission regarding the validity of utility EM&V programs and annual filings and to suggest ongoing improvements to EM&V activities."

Staff, after reviewing input from the utilities and other stakeholders, will be responsible for selection of the IEM. The Commission may ultimately resolve any disputes as to the selection and retention of the IEM. The IEM's fees and expenses shall be paid by the utilities, and these costs will be included in EM&V budgets and cost-effectiveness calculations, and recovered through its EECR rider. Each utility shall pay its share of the costs based on a ratio of its number of customers to the total of customers for all utilities combined.

The fundamental role of the IEM is to be advisory in nature. In this role, the IEM is tasked with providing technical consultation services to participating Arkansas utilities, staff and interveners regarding strategies that will result in program compliance with EM&V rules or protocols approved by the Commission. In the interest of regulatory economy, the IEM will work with staff, utilities and other parties to each utility's EE docket to ensure that any EM&V advisory input reflects the informed consideration of all parties.

To ensure fully independent evaluations, the IEM will be engaged throughout the ongoing process of prioritizing evaluation activities and budgets, defining evaluation objectives and methodologies, reviewing evaluation results, and a range of other continuing and related EM&V activities. At the conclusion of such an annual a process, the IEM will have the responsibility of issuing a report to the Commission regarding the integrity of utility EM&V programs and activities, and to suggest ongoing improvements to EM&V activities. After the issuance of the report and a reasonable comment period, the Commission may use the report as a basis to issue orders to the utilities.

The IEM will provide advisory assistance to the utilities, Staff and the interveners, including but not limited to the following broad categories. This list is intended to be illustrative.

- Assure of compliance with national EM&V "best practices," Commission-approved protocols and the TRM.
- Verify credentials, performance and independence of EM&V contractors and vendors.
- Assure adequacy of individual utility EM&V program budgets and the timing and prioritization of evaluation projects.
- Review utility energy efficiency program evaluation projects and EM&V methodologies, including PWC administered multi-utility evaluations.

- Recommend improvements to EM&V processes and procedures, including those related to custom program projects.
- Recommend improvements to the overall EM&V decision-making process and each utility's program design(s) and program implementation as it relates to EM&V.
- Assure timely updates and/or expansion of deemed savings and the TRM.
- Recommend updates to DSM potential and baseline studies as well as recommend appropriate multi-utility PWC EM&V efforts.
- Recommend additional financial (or other) resources that may be necessary for the effective functioning of the IEM or EM&V stakeholder process.

PROTOCOL J: BEHAVIOR-BASED PROGRAM EVALUATION PROTOCOL

This protocol was developed to reflect the evolving nature of energy efficiency programs now offered by utilities and third-party administrators in Arkansas. This protocol specifically addresses the prescribed approach to conducting evaluations for behavior-based programs. This protocol was developed based upon the recommended best practices described in the State and Local Energy Efficiency (SEE) Action Network (published by Laurence Berkeley National Laboratory (LBNL) report, "*Evaluation, Measurement, and Verification (EM&V) of Residential Behavior-Based Energy Efficiency Programs: Issues and Recommendations*." [8] The IEM has extracted the appropriate protocol specifications after receiving permission from the authors and the SEE Action Network. Furthermore, this protocol cites the sections in which these specifications are referenced.

Behavior-based energy efficiency programs are those that use strategies intended to affect consumer energy use behaviors to achieve energy and/or peak demand savings. Program types typically include outreach, education, competition, rewards, benchmarking and/or feedback elements. Such programs may rely on changes to consumers' *habitual* behaviors (e.g., turning off lights) or *one-time* behaviors (e.g., changing thermostat settings). In addition, these programs may target purchasing behaviors (e.g., purchases of energy efficient products or services), often in combination with other programs (e.g., rebate programs or direct install programs).

Key Definitions

Conducting evaluations of these programs first requires defining the following key terms. These definitions are cited from the SEE Action/LBNL Report 2012:

- **Treatment Group:** the group of households that are assigned to receive the treatment
- **Control Group:** the group of households that are assigned not to receive the program
- **Experimental Design:** a method of controlling the way that a program is designed and evaluated in order to observe outcomes and infer whether or not the outcomes are caused by the program
- **Randomized Controlled Trial (RCT):** a type of experimental design; a method of program evaluation in which households in a given population are randomly assigned into two groups — a treatment group and a control group — and the outcomes for these two groups are compared, resulting in unbiased program savings estimates

Quasi-Experimental Design: a method of program evaluation in which a treatment group and a control group are defined but households are not randomly assigned to these two groups, resulting in program savings estimates that may be biased (LBNL Report 2012, p. 14).

[8] State and Local Energy Efficiency Action Network. 2012. *Evaluation, Measurement, and Verification (EM&V) of Residential Behavior-Based Energy Efficiency Programs: Issues and Recommendations*. Prepared by A. Todd, E. Stuart, S. Schiller, and C. Goldman, Lawrence Berkeley National Laboratory. http://behavioranalytics.lbl.gov.

The protocol begins with a detailed discussion regarding the recommended evaluation design, then continues with a discussion regarding model specification, and concludes with other evaluation issues. The protocol also include models that are appropriate for evaluators to use when determining the likely effects pre- and post- of behavior-based programs.

EVALUATION DESIGN

Since behavior-based programs may not always specify or track particular actions that result in energy savings, the recommended approach to determine the effects of behavior-based efficiency programs is randomized controlled trial (RCT), which will result in the most robust, unbiased program savings impact estimates. As an alternative, the protocols also allow for two quasi-experimental designs.

The Randomized Controlled Trials (RCT) Method

In an RCT, first a *study population* is defined and then the study population is *randomly assigned* to either the treatment or control group. Energy use data must be collected for all households in the treatment and control group in order to estimate energy savings. The estimate of energy savings is then calculated by comparing the difference between the measured energy usage of the treatment households relative to the energy usage of the control households. Measured energy use typically comes from utility meter data, often in monthly increments.

Random assignment means that each household in the study population is randomly assigned to either the control group or the treatment group based on a random probability, as opposed to being assigned to one group or the other based on some characteristic of the household (e.g., location, energy use, or willingness to sign up for the program).

Randomization eliminates pre-existing differences that are both observable differences (e.g., energy use or floor area of households) as well as differences that are typically unobservable (e.g., attitudes regarding energy conservation, number of occupants, expected future energy use, and occupant age) unless surveyed. Thus, because of this random assignment, an RCT control group is an ideal comparison group: it is statistically identical to the treatment group in that there are no pre-existing differences between the two groups, which means that selection bias is eliminated (LBNL Report, p. 24).

RCTs also eliminates the free rider concern during the study period because the treatment and control groups each contain the same number of free riders through the process of random assignment to the treatment or control groups. When the two groups are compared, the energy savings from the free riders in the control group cancel out the energy savings from the free riders in the treatment group. Furthermore, *participant spillover* is also automatically captured by an RCT design for energy use that is measured within a household. The resulting estimate of program energy savings, therefore, is an unbiased estimate of the net savings caused by the program (the true program savings) (LBNL Report, p. 24).

There are three basic program enrollment options for behavior-based programs: opt-out enrollment, opt-in enrollment, and encouragement design, which does not restrict or withhold participation in the program to any household. Program designs with *any* of these enrollment

options may use RCTs for evaluation and thus each enrollment option can yield unbiased savings estimates. With any of the enrollment options, the random assignment of households into treatment and control groups is the crucial step. All data from the randomization point forward should be analyzed to ensure internal validity.

When implementing RCTs with any of these three enrollment options, the first step is to define the target market and the eligible households that are included in the study population (i.e., the screening criteria). Often this screening process restricts the study population to specific geographies (zip codes or service areas), specific demographics (hard-to-reach, medical needs, elderly), specific customer characteristics (high energy users, dual fuel use, length of customer bill history), and specific data requirements (one year of historical energy data available, census information is available, smart meter installed). Another way to reduce the size of the study population is to randomly select households out of a larger population in order to form a smaller subset of households (LBNL Report 2012, p. 26).

Randomized Controlled Trials with Opt-Out Enrollment

In some cases, program administrators may want to enroll households using an opt-out method (see Figure 8). For this type of program, an RCT with an opt-out recruitment strategy determines that after households have been evaluated for program eligibility (i.e., *screened in* or *screened out*); the remaining households are placed in the study population and are randomly assigned to either the control group or the treatment group. The treatment group receives the program (but are allowed to opt out), and the control group does not receive the program (and are not allowed to opt in). Energy use data must be collected for all of the households in the control and treatment group, whether or not they opt out, in order to estimate energy savings without bias.

Randomized Controlled Trial? With Opt-Out Enrollment

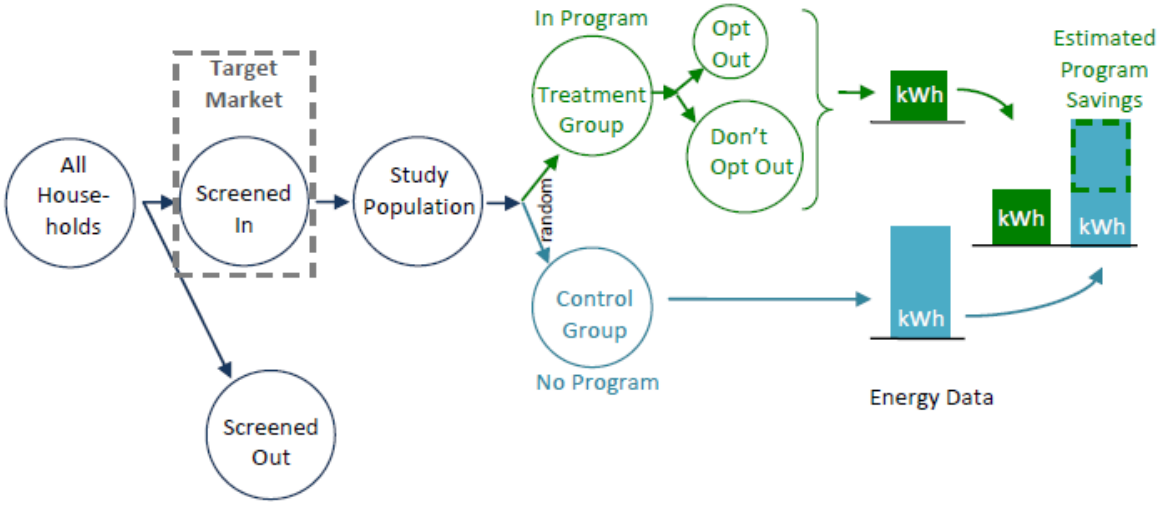


Figure 8: Randomized Controlled Trials with Opt-Out Enrollment[9]

[9] SEE/LBNL, page 14.

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

Randomized Controlled Trials with Opt-In Enrollment

For some types of behavior-based efficiency programs, an opt-in enrollment method may be more desirable than an opt-out method. Using an RCT with an opt-in recruitment strategy, after households have been screened, the program is marketed to the remaining households. These households decide whether they want to opt in to the program. The households that opt in define the group of households in the study population, which are then randomly assigned to either the control group or the treatment group. It is important that a randomly selected group of opt-in households is placed in the control group in order to have unbiased results: if households that opt in are compared with a control group of households that did not opt in, then these two groups contain very different types of households, which can result in selection bias and potentially invalid results.

There are two methods for randomizing the opt-in households into a treatment and control group: *recruit-and-delay* and *recruit-and-deny*. In *recruit-and-delay (*also called *waitlist) design*, households that opt in are told that the program is currently oversubscribed and some households may randomly be placed on a waitlist for a short time. In a *recruit-and-deny design*, households that opt in are told that the program is oversubscribed and so some households will be randomly chosen to participate. Energy use data must be collected for all households in the treatment and control group in order to estimate energy savings.

Randomized Controlled Trials with Encouragement Design

Often, program implementers want to allow households to opt in and do not want to deny or delay enrollment in the program. In this case, an *RCT with encouragement design* (sometimes called RED) yields an unbiased estimate and does not exclude *anyone* from participating in the program. However, an RED design typically involves a much larger sample size requirement to produce robust estimates of savings. An encouragement design is used after households have been screened by the utility; the households that meet the screening criteria define the study population and are randomly assigned to the control or treatment group. The treatment group is then encouraged to participate in the program. Some households may decide to opt in to the program while others may not. Households in the control group are not encouraged to participate, although, because the program is open to anyone, some of these households may learn of the program and decide to opt in. In order to have an unbiased estimate of energy savings, energy use data must be collected for all households in the treatment and control group for both the households that opted in to the program as well as those that did not.

Quasi-Experimental Methods

As discussed in the SEE/LBNL report, most quasi-experimental methods introduce potential bias into the savings estimates. For this reason the Arkansas Behavior-Based Evaluation Protocol strongly recommends use of one of the RCT approaches wherever possible.

There are situations, however, where use of an RCT can be extremely difficult, and thus as a secondary approach these protocols do allow the use of quasi-experimental methods if necessary. The recommended approaches are identified in the SEE/LBNL report as the Regression Discontinuity Method and the Variation in Adoption (With a Test of Assumptions) methods. Rather than recruit-and-delay or recruit-and-deny, which can be difficult to implement, both of these methods allow customers to opt in at the time of interest.

Regression Discontinuity Method

Among the quasi-experimental methods, regression discontinuity typically yields the most unbiased estimate of energy savings. However, it is also the most complicated method; it requires knowledge of econometric models and often requires field conditions that allow the evaluator to utilize this analytic technique, and is therefore not always practical. This method works if the eligibility requirement for households to participate in a program is a cutoff value of a characteristic that varies within the population. For example, households at or above a cutoff energy consumption value of 900 kWh per month might be eligible to participate in a behavior-based efficiency program, while those below 900 kWh are ineligible. In this case, the households that are just below 900 kWh per month are probably very similar to those that are just above 900 kWh per month. Thus, the idea is to use a group of households right below the usage cutoff level as the control group and compare changes in their energy use to households right above the usage cutoff level as the treatment group. This method assumes that the program impact is constant over all ranges of the eligibility requirement variable that are used in the estimation (e.g., that the impact is the same for households at all levels of energy usage), although there are

more complex methods that can be used if this assumption is not true.53 In addition, regression discontinuity relies on the eligibility requirement being strictly enforced.

Variation in Adoption (With a Test of Assumptions)

Under the Variation in Adoption method, also commonly referred to as a "rolling" control group design, the control group is made up of participants who sign up towards the end (or even following) the post-treatment period of interest.[10] The assumption is that the customers in the control group are similar to those in the treatment group, the primary difference being that they signed up at a slightly later period than the treatment group, thus their pre-participation period represents participant post-period usage in absence of the program.

As an example, assume that the year of interest for analysis is 2011, examining savings for all customers that signed up for the program (opted in) in 2011. The analysis would then need billing data for 2010-2012 to ensure that every participant has at least 12 months pre- and post-billing data. The control group, however, would be made up of customers that signed up in late 2012 (e.g., the last quarter) and early 2013 (e.g., the first and possibly second quarters). Any pre-program months for 2010-2012 for the control group can then be used in the billing analysis to represent the potential change from pre- to post-treatment of the participant group.

To implement this approach, the evaluation also needs to test the assumption that the only difference is the timing of enrollment, rather than any observable household characteristics. This assumption can be tested through a duration analysis or propensity score matching, as discussed in the SEE/LBNL document.

Recommended Approach for Analysis Model Specification

This section presents a number of important issues regarding the specification of the model.

Length of Study and Baseline Period

Relatively longer study periods and baseline data periods are likely to lead to greater precision of the estimated program impact. It is important to collect at least one full year of historical energy use data in order to have baseline data for each month and season since patterns of household energy use often vary by season. Thus, it is strongly advised that at least one full year (the twelve continuous months immediately prior to the program start date) of historical energy use data be available for each customer — both for those in the treatment group and in the control group — so that the baseline energy use reflects seasonal effects.

[10] Note this approach has been used by evaluators for many years for identifying a control group for opt-in programs such as low-income or residential audit programs.

Model Specification

Panel vs. aggregated data. Analysis models can either use energy data that are aggregated across time for both the pre- and post-program periods (e.g., average energy use for the period prior to and during the program year) or panel data (also called time series of cross-sectional data), which typically are data from multiple time points for pre- and post-program periods (e.g., monthly energy use for each month of the program and for the twelve months preceding the program). While the protocols allow for both, the panel data models are preferred because they result in a more precise estimate of energy savings, including seasonal variation of savings.

Comparison of energy usage vs. change in energy usage. Analysis models can be specified to estimate program savings by either comparing the energy saved by the treatment group (i.e., the change in energy use prior to and during the program) to the energy saved by the control group, or comparing the energy use of the treatment group to the control group during the program. The protocols recommend the difference between the change (i.e., the "difference of the differences") because it will typically be more precise due to the fact that the amount that the change in energy usage varies between households often varies less than the amount that energy usage varies between households.

Inclusion of control/interaction variables. All models must include variables that indicate whether a data point was taken prior to or during the program and whether each household or group of households is in the treatment group or the control group. In addition to these variables, models may include control variables (that may help explain the patterns of energy use unrelated to the program) and interaction variables (that provide insights as to the relationships between the program and other factors). The protocols strongly recommend the use of control variables (e.g., a variable that calculates the effect of a winter month or of weather on energy use), and recommend that interaction variables not be used for the primary savings model as they can potentially bias results.

Equivalency Check

Because the degree to which a savings estimate is unbiased depends on how similar the control group is to the treatment group, an important part of the analysis is validating that the two groups are equivalent. The correct procedure is to determine if the households in the treatment group have characteristics that are statistically similar to those in the control group. At a minimum, the evaluators should compare the monthly or yearly pre-program energy use and the distribution of pre-program energy use between the treatment and control group. Other possible covariates, which may or may not be available, include geographic location, dwelling characteristics (e.g., square footage), demographic characteristics (e.g., age, income), and psychographic characteristics (e.g., opinions) and any other baseline covariates for which data is available. This should be done whether the program is designed as an RCT or a quasi-experiment.

Statistical Significance

Evaluations of behavioral programs in Arkansas should follow the SEE/LBNL report recommendation that a null hypothesis (i.e., a required threshold such as the level or percentage of energy savings needed for the benefits of the program to be considered cost-effective) should be established, and the program savings estimate should be considered acceptable (and the null hypothesis should be rejected) if the estimate is statistically significant at the 5% level or lower. This threshold, which is greater (more precise) than that typically used by energy program evaluation (which is usually a confidence/precision level of 90/10), is the acceptable standard in behavioral sciences research. Given that behavioral programs are still a nascent approach to energy savings and cannot be measured through engineering approaches, this increased threshold appears acceptable. Note that evaluations of programs with smaller sample sizes (e.g., only hundreds of participants) should also attempt to assess savings at this significance level by using additional control variables as needed, but may report and claim savings at 90% significance.

Treatment of Households that Opt Out or Close Accounts

Households that opt out should never be excluded from the dataset; they should be included as part of the treatment group to avoid selection bias. In order to calculate an unbiased estimate of the effect of the program on those that did not opt out, the program impact should first be estimated including the opt-out households as part of the treatment group. Then, the estimate should be divided by the percentage of households that did not opt out. For example, if the estimate of the overall savings is two percent for all households in the treatment group, but 20 percent of the households opted out of the program, then the estimate of the effect of the program on the *treated* households that do not opt out is higher (i.e., 2.5%, which is 2% divided by 80%). This adjusted savings estimate should then be multiplied by the total number of households that did not opt out (or equivalent to multiplying the unadjusted savings by the total number of households, including those that opted out).

Households that close their accounts — including any changes in tenants at the same site — should be dropped entirely from the evaluation dataset (i.e., every data point for these households should be deleted) for both the control and treatment groups. However, there may be situations in which dropping households that closed accounts leads to biased estimates (e.g., younger and more mobile populations may be more responsive to behavior-based programs and may also be more likely to close accounts). In this case, if the analysis is done correctly with an indication that a specific sub-group of the population closed accounts, it may be better to include households that closed accounts.

Additional Evaluation Issues

Controlling for Double Counting

For programs in which efficiency measures can be tracked to a specific household (e.g., installation of insulation by a contractor), and sample sizes between the treatment and control group are equal, double-counted savings can be directly determined as the incremental participation in the non-behavior program (as shown in Figure 9 and in LBNL Report 2012, p.44). For cases where the sample sizes between the treatment and control group are not equal, the double counting should be quantified on the basis of the difference in per-participant savings from non-behavior programs.

Because the program and evaluation design of the behavior-based program utilizes a treatment and control group, we can infer that the double-counted savings were caused by the behavior-based program. While the SEE/LBNL report stops short of recommending how to handle the assignment of savings to program, the report does suggest that it is reasonable to assign at least half of the double-counted savings to the behavior-based efficiency program. The protocols agree with this assessment and approve of assigning half of the double-counted savings to each program. Note, however, that this also means appropriately dividing the program costs, as well as adjusting the lifetime savings for the measures that are being assigned to the behavioral program (i.e., greater than one year). Due to the potential complexity of dividing the incremental double-counted savings, however, the protocols also approve assigning all the incremental double counted savings to the other (non-behavioral) program, as has been the standard evaluation practice at the time of development of these protocols.

Example of Double-Counted Savings

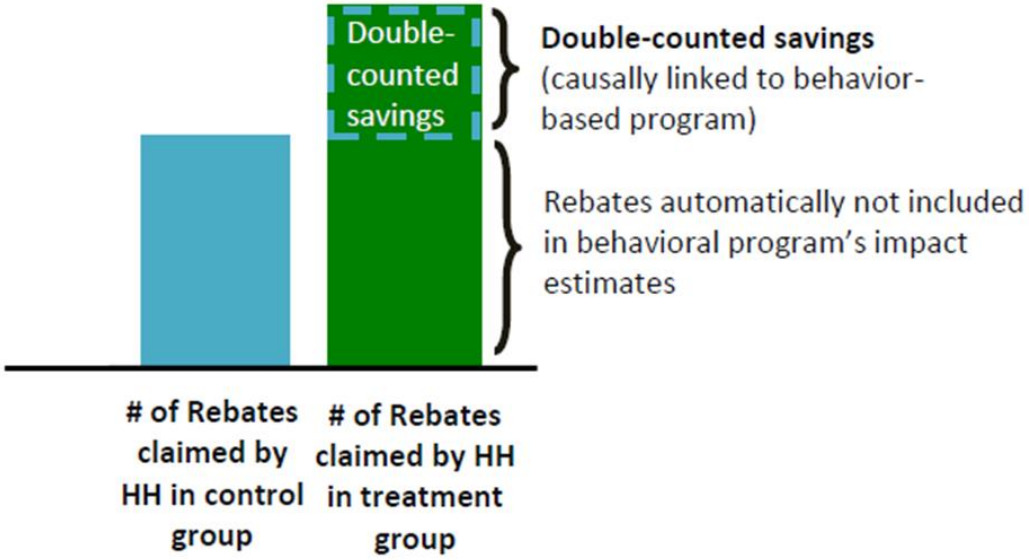


Figure 9: Example of Double-Counted Savings

Persistence of Program Savings

A control group should be maintained for every year in which program impacts are estimated, and the program is evaluated *ex post* every year initially and every few years after the program has been running for several years. Although an industry best practice has not been developed, this protocol encourages evaluators to not only assess the magnitude of savings each year, but also the actions that participants have taken to achieve measured savings.

Applying Savings to a New Population of Participants in Future Years

The protocols do not allow for directly applying program savings impact estimates from an initial program to an expanded program with a new population in a future year. Consistent with the SEE/LBNL report, the IEM recommends creating and maintaining a new control and treatment group, and evaluating the expanded program using the similar methods described in those protocols.

Reporting Requirements

The evaluation reports of behavioral programs shall contain, at a minimum, the following:

1. Description of which evaluation method was selected, and why, including details regarding how the control group was selected
2. Relevant sample sizes of the treatment and control groups, and number of months (or alternative time periods) pre/post that were included
3. Description of the final model specification, including which parameters were included
4. Full regression output for the model, including:
    a. For each parameter, the estimated coefficient and standard error
    b. Adjusted R^2
    c. Tests of joint significance of the model (e.g., F Test)
    d. Regression diagnostics (e.g., tests for multicollinearity, heteroscedasticity)
5. Assessment of quantity and treatment of overlap with other rebate programs (i.e., double counting)
6. A 95% confidence interval for the monthly or annual savings
7. Any surveys or other primary data collection that was conducted to assess how the savings were achieved

Summary of Acceptable Model Specifications for Behavior-Based Programs

Panel Data Model with Fixed Effects (Comparing Change in Energy Use)

Basic Model:

(1.1)

$$y_{it} = \mu_i + \underbrace{\beta Treat_{it}}_{\text{Variable indicating whether household i is in the program at time t}} + \underbrace{\delta Post_t}_{\text{A time variable indicating pre or post treatment}} + \underbrace{\varepsilon_{it}}_{\text{Error}}$$

y_{it}: Energy use of household i at time t

μ_i: Household fixed effects

$\beta = $ Estimated Program Savings Impact

Basic Model + Month-of-Year Time Control Variables:

(1.2)

$$y_{it} = \mu_i + \underbrace{\beta Treat_{it}}_{\text{Variable indicating whether household i is in the program at time t}} + \underbrace{\delta_1 Post_t}_{\text{A time variable indicating pre or post treatment}} + \underbrace{\delta_2 MonthOfYear_t}_{\text{A time variable indicating the month of year (from 1 to 12)}} + \underbrace{\varepsilon_{it}}_{\text{Error}}$$

y_{it}: Energy use of household i at time t

μ_i: Household fixed effects

$\beta = $ Estimated Program Savings Impact

Basic Model + Month-of-Sample Time Control Variables:

(1.3)

$$y_{it} = \mu_i + \underbrace{\beta Treat_{it}}_{\text{Variable indicating whether household i is in the program at time t}} + \underbrace{\delta MonthOfSample_t}_{\text{A time variable indicating the month of the data sample, from 1 to the last month in the sampmle (e.g., 1 to 36)}} + \underbrace{\varepsilon_{it}}_{\text{Error}}$$

y_{it}: Energy use of household i at time t

μ_i: Household fixed effects

$\beta = $ Estimated Program Savings Impact

Basic Model + Month-of-Sample Time Control Variables + Other Control Variables:

(1.4)

$$y_{it} = \mu_i + \underbrace{\beta Treat_{it}}_{\text{Variable indicating whether household i is in the program at time t}} + \underbrace{\delta MonthOfSample_t}_{\text{A time variable indicating the month of the data sample, from 1 to the last month in the sampmle (e.g., 1 to 36)}} + \underbrace{\gamma Weather_{it}}_{\text{Other control variables, e.g., HDD, CDD}} + \underbrace{\varepsilon_{it}}_{\text{Error}}$$

y_{it}: Energy use of household i at time t

μ_i: Household fixed effects

$\beta = $ Estimated Program Savings Impact

Basic Model + Month-of-Sample Time Control Variables + Interaction Variables (Unacceptable for estimating program impact):

$$y_{it} = \mu_i + \underbrace{\beta Treat_{it}}_{\text{Variable indicating whether household i is in the program at time t}} + \underbrace{\delta MonthOfSample_t}_{\text{A time variable indicating the month of the data sample, from 1 to the last month in the sample (e.g., 1 to 36)}} + \underbrace{\gamma Treat_{it} Weather_{it}}_{\text{Other control variables, e.g., HDD, CDD}} + \underbrace{\varepsilon_{it}}_{\text{Error}} \quad (1.5)$$

Energy use of household i at time t

Household fixed effects

β Does NOT equal Estimated Program Savings Impact

Time-Aggregated Data Model (Comparing Change in Use)

Basic Model (Difference-in-Differences)[11]:

(1.6)

$$\underbrace{s}_{\text{Estimated Program Energy Savings Impact}} = \underbrace{\left(\underbrace{\sum_{i \in Treatment\ group} \sum_{t \in pre\ program} y_{it}}_{\text{Average energy use of all households i that are in the treatment group, during the period t before the household was enrolled in the program}} - \underbrace{\sum_{i \in Treatment\ group} \sum_{t \in post\ program} y_{it}}_{\text{Average energy use of all households i that are in the treatment group, during the period t after the household was enrolled in the program}} \right)}_{\text{Average change in energy use (average energy saved) for the treatment group}} - \underbrace{\left(\underbrace{\sum_{i \in Control\ group} \sum_{t \in pre\ program} y_{it}}_{\text{Average energy use of all households i that are in the control group, during the period t before the household was enrolled in the program}} - \underbrace{\sum_{i \in Control\ group} \sum_{t \in post\ program} y_{it}}_{\text{Average energy use of all households i that are in the control group, during the period t after the household was enrolled in the program}} \right)}_{\text{Average change in energy use (average energy saved) for the control group}}$$

[11] Note that this method relies on the assumption that the program begins for every control and treatment household that is being analyzed at the same time (average energy saved) the same time. For example, if the treatment coincides with the billing cycle and billing cycles are different for different households, then time-aggregated models should not be used; instead, panel data models should be used.

Basic Model + Control Variables:

Adding control variables to the above *difference-in-difference* model can be written in several ways; the treatment averag**e** en**e**rgy sav**e**d) and control variables could be listed in a table, or an equation could be written similar to those in the panel data for the treatment group model section. The key distinction is that the energy variable is aggregated over time to get one observation per household before the program began, and one observation per household after the program began.

Panel Data Model without Fixed Effects (Comparing Use Rather Than Change in Use)

Basic Model:

(1.7)

$$y_{it} = \underbrace{\beta Treat_i}_{\text{Variable indicating whether household i is in the program (is in the treatment group) or is in the control group}} + \underbrace{\varepsilon_{it}}_{\text{Error}}$$

Energy use of household i at time t

β = Estimated Program Savings Impact

Basic Model + Month-of-Year Time Control Variables:

(1.8)

$$y_{it} = \underbrace{\beta Treat_i}_{\text{Variable indicating whether household i is in the program (is in the treatment group) or is in the control group}} + \underbrace{\delta_2 MonthOfYear_t}_{\text{A time variable indicating the month of year (from 1 to 12)}} + \underbrace{\varepsilon_{it}}_{\text{Error}}$$

Energy use of household i at time t

β = Estimated Program Savings Impact

Basic Model + Month-of-Sample Time Control Variables:

(1.9)

$$y_{it} = \underbrace{\beta Treat_i}_{\text{Variable indicating whether household i is in the program (is in the treatment group) or is in the control group}} + \underbrace{\delta MonthOfSample_t}_{\text{A time variable indicating the month of the data sample, from 1 to the last month in the sampmle (e.g., 1 to 36)}} + \underbrace{\varepsilon_{it}}_{\text{Error}}$$

Energy use of household i at time t

β = Estimated Program Savings Impact

Basic Model + Month-of-Sample Time Control Variables + Other Control Variables:

$$y_{it} = \underbrace{\beta Treat_i}_{\text{Variable indicating whether household i is in the program (is in the treatment group) or is in the control group}} + \underbrace{\delta MonthOfSample_t}_{\text{A time variable indicating the month of the data sample, from 1 to the last month in the sampmle (e.g., 1 to 36)}} + \underbrace{\gamma Weather_{it}}_{\text{Other control variables, e.g., HDD, CDD}} + \underbrace{\varepsilon_{it}}_{\text{Error}} \quad (1.10)$$

Energy use of household i at time t

$$\beta = \text{Estimated Program Savings Impact}$$

Basic Model + Month-of-Sample Time Control Variables + Interaction Variables (Unacceptable for estimating program impact):

$$y_{it} = \mu_i + \underbrace{\beta Treat_i}_{\text{Variable indicating whether household i is in the program (is in the treatment group) or is in the control group}} + \underbrace{\delta MonthOfSample_t}_{\text{A time variable indicating the month of the data sample, from 1 to the last month in the sample (e.g., 1 to 36)}} + \underbrace{\gamma Treat_{it} Weather_{it}}_{\text{Other control variables, e.g., HDD, CDD}} + \underbrace{\varepsilon_{it}}_{\text{Error}} \quad (1.11)$$

Energy use of household i at time t; Household fixed effects

$$\beta \text{ Does NOT equal Estimated Program Savings Impact}$$

Time-Aggregated Data Model (Comparing Use)

Basic Model (Difference-in-Averages):

$$\text{Estimated Program Energy Savings Impact} = \underbrace{\sum_{i \in Control\ group} \sum_{t \in post\ program} y_{it}}_{\text{Average energy use of all households i that are in the control group, during the period t after the household was enrolled in the program}} - \underbrace{\sum_{i \in Treatment\ group} \sum_{t \in post\ program} y_{it}}_{\text{Average energy use of all households i that are in the treatment group, during the period t after the household was enrolled in the program}} \quad (1.12)$$

Basic Model + Control Variables:

Adding control variables to the above *difference-in-averages* model can be written in several ways; the treatment and control variables could be listed in a table, or an equation could be written similar to those in the panel data model section. The key distinction is that the energy variable is aggregated over time to get one observation per household after the program began.

All analyses should be performed with the untransformed variable of energy use, rather than the transformed variable of the natural log of energy use (typically written *ln(y)*). Analyzing the data with an untransformed variable will yield an estimate of the effect of the program averaged over all households in kilowatt-hour terms, while analyzing the data with the log of energy will yield the effect of the program averaged over all households in percentage terms. Because higher energy use households may tend to save more energy as a result of the program, performing the analysis with the log of energy use may lead to a misleadingly small estimate of energy use.

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

PROTOCOL K: LEAKAGE

Definition of Leakage

Cross-territory sales, or "leakage," occur when program-incented efficient products are installed outside of the funding utility's service territory. When this occurs, the energy and demand savings from the incentivized product are not being realized within the territory that paid for, and is claiming savings for, the unit. Upstream programs are particularly vulnerable to leakage as the rebate recipient is unknown and sales not restricted based on utility.[12]

While variations of leakage may occur in other forms (e.g., utility customers purchasing and installing a rebated product in a vacation home that sits outside the sponsoring service territory), these protocols are intended for upstream programs where the purchaser of the product is unknown. While the predominant upstream program is currently for efficient lighting, future upstream programs should also rely on these protocols. Other program types should handle cross-service territory installations through other means, such as telephone or on-site verification.

Program administrators should make concerted efforts to mitigate leakage. The most common effort is to limit participation of retail storefronts that are on the perimeter of the service territory. Protocol K provides a brief synopsis regarding the Commission Orders regarding leakage and then focuses on the methodologies for determining leakage, as required by the Commission Orders in Docket No. 07-082-TF, Order No. 63, and Docket No. 07-085-TF, Order No. 85.[13]

Policy Regarding Leakage

The Commission determined that leaked energy efficiency measures do not generate Lost Contribution to Fixed Costs (LCFC), therefore limiting LCFC collection to proven lost fixed costs.

The Commission, therefore, determined that energy efficiency program leakage shall be incorporated into LCFC calculations as determined by the Orders, based upon a leakage calculation methodology that will be approved in the 2013 TRM update process after further refinement by the PWC, and implemented through the 2014 annual EECR adjustment.

With regard to net benefits and utility energy efficiency performance incentives, the Commission determined that the state, and ratepayers as a whole, benefit from energy saved and reduced demand. Also, sponsoring utility ratepayers have an interest in the continuation and refinement of a program to speed the implementation of emerging technologies such as LED lighting.
Based on aggregate ratepayer and statewide benefits, the need to address emerging lighting and other technologies, the need to promote efficient program administration, and the orderly

[12] In upstream programs (also referred to as Point-of-Purchase Markdown Programs) the incentives go directly to the retailer or manufacturer rather than the customer.
[13] For more information on the policy implications of leakage see Arkansas Public Service Commission Docket No. 07-082-TF, Order No. 63, and Docket No. 07-085-TF, Order No. 85.

implementation of Commission guidance, the Commission determined that leakage should not be subtracted from energy efficiency goal achievement and net benefits calculations for PY 2012 or, absent a change in circumstances and further proceedings, for future years.

Research Recommendations to Address Leakage

Since retailers perceive customer contact information and sales data as highly confidential, estimating leakage is an inherently difficult task. Protocol K acknowledges this difficulty, with the understanding that any research approach will have at least some threats to validity. Protocol K therefore, recommends three research approaches, and further encourages evaluators to attempt as many of these as possible so as to minimize bias that may be introduced from any individual approach.[14]

However, if all three methods are not possible due to budget limitations, retailer refusal to participate, or other possible reasons, the methods are presented in order of preference from what is believed to be lowest to highest threats to validity. Combinations of approaches may also be used. For example, intercepts might be used only at the high volume stores to determine leakage, and geo-mapping could be used for the lower-volume retailers where conducting intercepts might be cost-prohibitive.

Research Method 1: Customer Intercept Surveys

Intercept surveys are the preferred data collection approach, as they gather primary data on specifically which customers purchase products at participating stores. Customer intercept surveys rely on in-store interviews with customers purchasing efficient products at participating stores. Likely leakage is then estimated by the number of products purchased for use outside of the service territory, divided by the total number of purchased products.

If using this approach, the evaluator needs to survey a geographically representative sample of participant retailers, and focus on retailers selling high volumes of incented bulbs. This ensures that the findings can be applied to the entirety of stores and distributed bulbs.

In recent years, retailers have been increasingly hesitant to allow evaluators to survey customers inside their stores. Many have restricted these types of activities entirely. Residential program implementers and utilities can add significant value by gaining permission for these surveys and should actively encourage participant retailers to permit third-party evaluators in their stores.

Potential bias from intercepts can occur in a number of ways. For example, retailer permission may be inconsistent; retailers may limit intercepts to certain days (e.g., during promotions); weekend vs. weekday customer traffic may differ. Protocol K therefore recommends that evaluators minimize this bias by attempting to include as many participating retailers as possible, not limiting the days of the intercepts to in-store promotions, and including a mix of weekend and weekdays.

[14] The IEM acknowledges that the approaches presented here do not represent an exhaustive list of all options. Alternative methods, or modifications to these methods, are allowed under Protocol K with prior approval from the IEM team.

One acknowledged challenge of intercepts is to include a mix of all retailer types, particularly for retailers that sell few program products, as it may not be cost-effective to have an in-store surveyor on-site conducting the interviews. Protocol K recommends two potential strategies to deal with potential retailer/distribution channel bias:

- ***Include sales of efficient and standard efficiency products.*** The evaluator may expand the survey to include a mix of efficient vs. standard efficiency products. For example, intercept surveys can include purchasers of program-incented products (e.g., CFLs), non-program high efficiency products, and standard efficiency products (e.g., incandescent or Energy Independence and Security Act (EISA) compliant halogen bulbs). All participants in the surveys can be asked about the intended location of the production installation, which can be used to derive the leakage estimate. This assumes that the leakage of non-program efficient products and standard efficiency products matches that of program-incented products. Because of this assumption, this method is only recommended for retailers where the purchase of inefficient products is so limited as to make the intercept approach cost-prohibitive.

- ***Weighting results to the population of retailers.*** Another approach, consistent with the recommendation from the Uniform Methods Project, is to weight the results from the sample population to the universe of participating retailers.[15] For those distribution channels that have not received intercept surveys, the evaluator should first assess how the leakage might differ and then apply extrapolated values.

As an example, if intercepts are only conducted at retailers that represent 75 percent of all program sales, the determined leakage rate should reflect the 75 percent of sales. For the remaining 25 percent of sales, the leakage rate can be adjusted based on factors such as proximity to the service territory border, distribution channel (e.g., groceries might have lower leakage rates than large home improvement stores), geo-mapping (as described below) and any other factors that can be determined from the actual in-store intercepts that were conducted.

Finally, in-store intercepts offer another important advantage because they can be used to determine the percentage of incented products that have been sold to commercial customers. Commercial customers typically use lighting products for more hours per day than do residential customers, and typically have higher peak coincidence factors, thus leading to higher savings. Savings from incented products sold to the sponsoring program administrator's commercial customers can be claimed towards both program goals and cost-recovery.[16]

[15] See http://www1.eere.energy.gov/wip/pdfs/53827-6.pdf for a final copy of the Uniform Methods Project Residential Lighting Evaluation Protocols.

[16] Note, however, that if sales to commercial customers are going to be determined, the survey should also probe for type of business, as hours of use can differ significantly between business types.

Research Method #2: Geo-Mapping and General Population Surveys

The evaluations of the 2012 upstream lighting programs relied on an approach of "geo-mapping", whereby each program storefront is assigned a leakage score based on the percentage of the sponsoring utilities customers (vs. other utility customers) that lie within a 60-minute drive time from that store. The 60-minute drive time radius, however, was adjusted for stores where an alternative storefront in the same distribution channel existed (i.e., with the assumption that customers will drive to the closest grocery, large home improvement, discount store, or any other store type within the same distribution channel).

The primary strength of this approach is that each storefront is assigned its own leakage score, and the overall estimate of leakage represents a weighted average of program storefronts. The primary limitation of this approach, however, is that the drive time/shopping estimates will all be based on evaluator judgment in the absence of any actual primary data collection.

Protocol K specifies that the geo-mapping approach will be supplemented with telephone surveys to test to the assumptions regarding drive time. The specific steps would include:[17]

1. ***Overlay utility service areas and population data from the U.S. Census Bureau***. This step matches the utility to all population points within its service area based on the census data in the highest publically available resolution available.

2. ***Estimate the customer base for each retail store by calculating store territory based on drive time to the nearest store***. In this step, territories for each program storefront are developed, looking at various increments of drive time. The drive times between adjacent storefront distribution channels are split in the middle between the two stores. So if two grocery stores are located 30 minutes apart, the analysis should assume that customers drive to the closest grocery store (i.e., no more than 15 minutes for those customers that sit in between the two stores).

3. ***Conduct a telephone survey to identify customer-shopping patterns for incented products***. A randomized survey of residential customers in the sponsoring utility service territory, as well as the neighboring utility service territory/ies, should be conducted to test the drive-time assumptions in the leakage model. For example, the survey can identify customers that purchased an energy efficient product in the last year (program or non-program), and where they purchased the product from, as well as where they normally shop for the product. A survey such as this can leverage a general population survey for the sponsoring utility customers, supplementing the sample with a random digit dial (RDD) sample of neighboring utility customers.

4. ***Allocate subsidies for each store to the population within the store territory, using actual sales data***. Each store would then be assigned a leakage score, with some

[17] These steps align closely to the steps outlined in the Cadmus May 1, 2013 memo entitled "CFL Subsidy Leakage in Arkansas."

percentage of the incented product assigned to the sponsoring utility customers and some percentage assumed to have been sold to neighboring utility customers.

5. ***Calculate leakage by summarizing the subsidies received by the population in and out of the utility service areas for each store territory***. Combine all the stores to determine an overall leakage rate for the sponsoring utility. This leakage rate should also include error bounds that account for potential error in the mapping of the census data to the service territory.

Research Method #3: Opt-In Surveys

Using "opt-in" surveys is also an option to determine leakage. This approach offers several advantages: it can to reach all participating storefronts and all customers purchasing incented products, it can be relatively inexpensive to implement, it can provide cross-customer class (commercial) sales, and it will provide a leakage estimate based on actual customer data. The limitation of this approach, however, is that as an "opt-in" survey response rates are typically extremely low, and thus may lead to significant non-response bias.

Should this method be implemented, the Protocol K specifies that:

- ***The evaluation should try to include all incented products from all program storefronts***. Ideally, each product would have a label/note about how to participate in the study.

- ***Customers should receive an incentive for their participation in the study.*** This may include a combination of individual incentives (e.g., a $10 gift card for the next time they shop at that participating retailer) as well as entry into a drawing for a larger, higher profile prize.

- ***The survey provides a multimodal approach.*** These can include sending in a reply card, taking the survey on-line, or sending customer contact information via other means, such as texting (allowing a call back from the research firm).

- ***Derive stratified leakage rates that are then weighted to the population.*** Similar to the intercept and geo-mapping methods, the leakage estimate should attempt to differentiate leakage rates by distribution channel or storefront, then weighting the findings up to the total population.

Reporting Leakage

As a way to provide additional transparency for calculating leakage, the evaluator should include a leakage rate for each participant storefront, along with the method for used for determining the leakage rate. In addition, details regarding the assumptions should be included (e.g., if the geo-mapping method is used, the average drive time should be included).

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

Relationship of Leakage with Net-to-Gross

Each method presented here is specific to estimating leakage (i.e., the percentage of incented product sold to non-sponsoring utility customers). The survey questions, particularly in the intercept and opt-in surveys approaches, should be brief and focus on customer name, contact information, and expected location of installation (address, self-reported utility, customer sector, and – if applicable – business type). These questions are completely distinct from net-to-gross (NTG) analysis, where the survey questions are likely to examine customer intentions, price-elasticity, and other potential parameters to estimate the likelihood that the customer would have purchased the incented product in absence of the program rebate. The estimated leakage values and NTG values, therefore, should be incorporated as two separate adjustments to the savings estimates as required by the relevant Commission rules and Orders.

The IEM does acknowledge, however, that NTG approaches may provide an NTG estimate that includes sales to customers outside the sponsoring utility service territory. This is appropriate for calculation of energy savings for the assessment of energy efficiency goal achievement and net benefits calculations, as leakage is not deducted from these estimates, and therefore the net savings should be based on utility and non-utility customers.

For purposes of the LCFC, however, leakage is incorporated, and thus the NTG ratio should be based off of sponsoring-utility customers. This would require calculating two separate NTG ratios, one for sponsoring utility customers and one for non-sponsoring utility customers. While this is worthy of consideration, the calculation of dual NTG ratios will introduce additional uncertainty, and thus should be subject to budget and timeline considerations, as well as prior approval of the IEM Team. In absence of a sponsoring-utility specific NTG ratio, the overall NTG ratio (i.e., that which incorporates all purchasers) should be used for purposed of both the energy efficiency goal achievement and the LCFC calculations.

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

References

Barnes, H., Jordan, G. 2006. *EERE Guide for Managing General Program Evaluation Studies*, prepared for the Office of Energy Efficiency and Renewable Energy (EERE), Office of Planning, Budget and Analysis for Lockheed Martin Aspen Lead Contractor with Sandia National Laboratories, February.

Energy Efficiency Program Evaluation: A Guide to the Guides, 2007. National Action Plan for Energy Efficiency and the Department of Energy's Office of Energy Efficiency & Renewable Energy, Washington D.C.

International Performance Measurement & Verification Protocol: Concepts and Options for Determining Energy and Water Savings, 2010 Volume I, prepared by the International Performance Measurement& Verification Protocol Committee, www.evo-world.org.

Johnson, K., Cook. G. 2009. *Program Evaluation for Missouri Gas Energy's Water Heater Evaluation*, Missouri Gas Energy, Kansas City, MO. January.

_____, Cook, G., Merurice, J. 2010. *Calculating Net-to-Gross Analysis Memorandum*, EmPOWER Maryland Program Evaluation, Navigant, Chicago, IL. April.

_____, Cook, G., Cooney, K. 2011. *Process Evaluation for the Partners in Energy Savings (PIES) Program Collaborative, Volumes 1&2*, Atmos Energy, Colorado Natural Gas, Eastern Colorado Utilities & SourceGas, Lakewood, CO. March.

_____, & Eisenberg, G. 2011. *New York State Process Evaluation Protocols: A Supplement to the New York State Guidelines Update 2011*, New York State Evaluation Advisory Group, New York. November pp. 21-22.

National Action Plan for Energy Efficiency (NAPEE) Action Plan and Resource Guides for Process, Impact Evaluations and Understanding Cost-Effectiveness of Energy Efficiency Programs, 2007. DOE/EPA, Washington, D.C. November.

Northeast Energy Efficiency Partnerships, 2010. *Regional EM&V Methods and Savings Assumptions Guidelines*, the Regional Evaluation, Measurement & Verification Forum, May.

Reynolds, D., Johnson, K., & Cullen, G. 2009 *Best practices for developing cost-effective evaluation, measurement, and verification plans: lessons learned from 12 northern California municipal utilities*, European Council for an Energy Efficient Europe (ECEEE), France

Research Into Action, 2010. *Research Supporting an Update of BPA's Measurement and Verification Protocols*, prepared for Bonneville Power Administration, April.

Schiller, S. 2007. *Model Energy Efficiency Program Impact Evaluation Guide, National Action Plan for Energy Efficiency*,

_____, 2010. *Energy Efficiency Evaluation, Measurement & Verification: Issues & Opportunities*, Presentation to the Regulatory Assistance Project Training Advocates' Training,

TechMarket Works Team, 2006. *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals*, prepared for the California Public Utilities Commission. April.

Glossary

This glossary is drawn from three evaluation-related reference documents:

- 2007 IPMVP
- 2004 California Evaluation Framework
- 2006 DOE EERE Guide for Managing General Program Evaluation Studies

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

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

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

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

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

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

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

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

California Measurement Advisory Council (CALMAC): An informal committee made up of representatives from the California utilities, state agencies, and other interested parties. CALMAC provides a forum for the development, implementation, presentation, discussion, and review of regional and statewide market assessment and evaluation studies for California energy efficiency programs conducted by member organizations.

Coincident Demand: The metered demand of a device, circuit, or building that occurs at the same time as the peak demand of a utility's system load or at the same time as some other peak

of interest, such as building or facility peak demand. This should be expressed in a way that indicates the peak of interest (e.g., "demand coincident with the utility system peak"). Diversity factor is defined as the ratio of the sum of the demands of a group of users to their coincident maximum demand. Therefore, diversity factors are always equal to one or greater.

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

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

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

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

Cost-effectiveness: An indicator of the relative performance or economic attractiveness of any energy efficiency investment or practice. The present value of the estimated benefits produced by an energy efficiency program is compared to the estimated total costs to determine if the proposed investment or measure is desirable from a variety of perspectives (e.g., whether the estimated benefits exceed the estimated costs from a societal perspective).

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

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

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

Dependent Variable: One that changes and is affected by the independent variables. Examples include weather, energy usage, housing type or location.

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

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

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

Energy Efficiency Measure: Installation of equipment, subsystems or systems, or modification of equipment, subsystems, systems, or operations on the customer side of the meter, for the purpose of reducing energy and/or demand (and, hence, energy and/or demand costs) at a comparable level of service.

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

Error: Deviation of measurements from the true value.

Evaluation: The performance of studies and activities aimed at determining the effects of a program; any of a wide range of assessment activities associated with understanding or documenting program performance, assessing program or program-related markets and market operations; any of a wide range of evaluative efforts including assessing program-induced changes in energy efficiency markets, levels of demand or energy savings, and program cost-effectiveness.

Ex ante Savings Estimate: Forecasted savings used for program and portfolio planning purposes (from the Latin for "beforehand").

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

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

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

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

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

Independent Variables: The stand-alone factors that affect energy use and demand, but cannot be controlled (e.g., weather, occupancy, age, gender). Regression analysis tries to determine the relationship between dependent and independent variables.

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

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

Leakage: Cross-territory sales that occur when program-incented efficient products are installed outside of the funding utility's service territory.

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

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

Market Transformation: A reduction in market barriers resulting from a market intervention, as evidenced by a set of market effects, that lasts after the intervention has been withdrawn, reduced, or changed.

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

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

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

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

Monitoring: Gathering of relevant measurement data, including but not limited to energy-consumption data, over time to evaluate equipment or system performance. Examples include chiller electric demand, inlet evaporator temperature and flow, outlet evaporator temperature, condenser inlet temperature, and ambient dry-bulb temperature and relative humidity or wet-bulb temperature, for use in developing a chiller performance map (e.g., kW/ton vs. cooling load and vs. condenser inlet temperature).

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

Net-to-Gross Ratio (NTGR): A factor representing net program savings divided by gross program savings that is applied to gross program impacts, converting them into net program load impacts after adjustments for free ridership and spillover.

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

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

Participant: A consumer who received a service offered through the subject efficiency program in a given program year. The term "service" is used in this definition to suggest that the service can be a wide variety of services, including financial rebates, technical assistance, product installations, training, energy efficiency information or other services, items, or conditions. Each evaluation plan should define "participant" as it applies to the specific evaluation.

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

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

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

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

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

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

Process Evaluation: A systematic assessment of an energy efficiency program for the purposes of documenting program operations at the time of the examination, and identifying and recommending improvements to increase the program's efficiency or effectiveness for acquiring energy resources while maintaining high levels of participant satisfaction.

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

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

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

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

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

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

Resource Acquisition Program: Programs designed to directly achieve energy and/or demand savings, and possibly avoided emissions, through the installation of new equipment.

Retrofit Isolation: The savings measurement approach defined in IPMVP Options A and B, and ASHRAE Guideline 14, that determines energy or demand savings through the use of meters to isolate the energy flows for the system(s) under consideration.

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

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

Simple Engineering Modeling (SEM): A basic calibrated engineering model used to determine energy impacts in a low-cost way**.**

Snapback (aka "take-back" or "rebound"): The increase in overall energy consumption after the installation of the efficiency measure, due to increased usage.

Spillover: Reductions in energy consumption and/or demand caused by the presence of the energy efficiency program that exceed the program-related gross savings of the participants. There can be participant and/or non-participant spillover rates depending on the rate at which participants (and non-participants) adopt energy efficiency measures or take other types of efficiency actions on their own (i.e., without an incentive being offered).

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

Stipulated Values: See "deemed savings."

Takeback Effect: See "rebound effect."

Uncertainty: The range or interval of doubt surrounding a measured or calculated value within which the true value is expected to fall with some degree of confidence.

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

Index

activities, 2, 6, 8, 16
additionality, *See Glossary*
adjustment, 12, *See Glossary*
aerators, 33
allowances, *See Glossary*
Ancova, See Glossary
annual, 21
assessment boundary, *See Glossary*
attribution, 54
audits, 16, 33
barriers, 6, 25
baseline, 7, 9, 11, 36, *see Glossary*
baseline adjustment, 7
best practices, 7
bias, *see Glossary*
boilers, 33
boundary, 7
budget, 7, 14
building, 12
building characteristics, 6
calibrate, 36
Calmac, *see Glossary*
categories, 21
CFL, 35
change, 7
characteristics, 17, 33
C&I, commercial and industrial, 30
code, 19
coincident demand, *see Glossary*
coincident peak demand, 21
commercial, 33
commercial measures, 41
comparison group, *see Glossary*
conditional savings analysis, *see Glossary*
confidence, *see Glossary*
consumption, 11, 22, 35
contractor, 19
control group, 74, 78
controls, 36
cooling, 22
cost savings, 16
cost-effectiveness, 33, *see Glossary*
costs, 8
CSA, *see Glossary*
custom measures, 42
customer class, 17, 21, 25, 33, 51, 53, 62
residential, multi-family, small C&I, large C&I, institutional
data, 2, 7, 8, 21
data collection, 9, 11, 19, 57
database, 17
database for energy-efficient resources, 97
deemed savings, 33, *see Glossary*
DEER, see Glossary
definition, 10
demand, 7, 21, *see Glossary*
dependent variable, *see Glossary*
diagnostic process evaluations, *see PROTOCOL A*
direct emissions, *see Glossary*
direct-install, 33
double counting, 82
education, 8
EERE 2006, 5
Effective Useful Life, *see Glossary*
effectiveness, 6
efficiency, 6
effort, 33
EM&V, 1, 3, 33
encouragement, 76, 78
end-use, 22
energy efficiency, *see Glossary*
energy efficiency measure, *see Glossary*
engineering, 33
engineering model, *see Glossary*
enrollment, 76
error, see Glossary
evaluation. 6, 8, *see Glossary*
ex ante savings estimate, *see Glossary*
ex post evaluation estimated savings, *see Glossary*
experimental design, 74
facility, 36
factors, 6, 12, 22
faucet, 33
food service, 33
forecasting, 22
free driver, *see Glossary*
free rider, 54, *see Glossary*
frequency, 2
furnaces, 33
goals, 7
granularity, 10
gross, 7
gross energy savings, 13
gross savings, *see Glossary*
heating, 22
hours of operation, 35
impact, 7, 21

impact evaluation, 10, *see Glossary*
implementation, 6, 8
implementers, 10
independent, 16
independent variables, 12, *see Glossary*
indirect emissions, *see Glossary*
industrial, 33
inspections, 35
interactive factors, *see Glossary*
international performance measurement and verification protocol., SEE ipmvp
interviews, 7
Ipmvp, 13, 22, 33, 35
LCFC, 53
leakage, Protocol K, *see Glossary*
levels of effort. *see effort*
life cycle, 10
lighting, 36
load shapes, 22, *see Glossary*
M&V, see Glossary
market effect evaluation, *see Glossary*
market transformation, *see Glossary*
marketing, 19
measure, 6, 7, 18, 33
measure characteristics, *see characteristics*
measurement, *see Glossary*
measurement boundary, *see Glossary*
metering. see glossary
meters, 36
methodologies, 4
metrics, 2
monitoring, see Glossary
motor, 36
NAC, *See Normalized Annual Consumption Model*
NAPEE 2007, 5, 7, 15, 72, National Action Plan for Energy Efficiency
net, 7
net energy savings, 13
net savings, *see Glossary*
net-to-gross, 13, 24, 53, 54, 94, 95, 100
net-to-gross ratio, *see Glossary*
new construction, 35
non-participant, 6, 22, *see Gossary*
Normalized Annual Consumption, 22, *see Glossary*
NTG, 57
NTG Ratio, 55
NTGR, 24, 53
Objectives, 6
on-site inspections, 33
operations, 6, 25
opt-in, 76, 77
opt-out, 76
outcome, 6
outreach, 8, 19
paper review, 22
participants, 6, 22, *see Glossary*
peak demand, *see Glossary*
perform, 35
performance, 6
performance standard, 11
persistence, 10, *see Glossary*
pilot, 8
portfolio, 15, *see Glossary*
portfolio level evaluations, *see PROTOCOLS A, B, C, D, E*
post-implementation, 21
post-program, 22
post-retrofit, 22
potential, *see Glossary*
precision, 10, *see Glossary*
premise, 24
premise characteristics, see characteristics
pre-program, 22
pre-retrofit, 22
prescriptive, 33
primary effects, *see Glossary*
process, 6, 8
process evaluation, see Glossary
process evaluation structure, *see PROTOCOLS A&B*
program, 7, 18, *see Glossary*
program administrator, 15
program evaluation, 10
program type, 25, 29, 33
  appliance recycling, audit, home performance, education, outreach, faucet, financing, pilot, rebate,
programs not meeting targets. *see PROTOCOL A*
project, *see Glossary*
project-specific, 11
protocol, 1, 2, 3, 17, 21
Protocol A, 25
Protocol B, 21
Protocol B1, 22
Protocol B2, 22
Protocol B3, 22
Protocol B4, 24
Protocol C, 25
Protocol C1, 26
Protocol C2, 29
Protocol D, 33
Protocol D1, 35
Protocol D2, 35
Protocol E , 51
Protocol E1, 52
Protocol E2, 52
Protocol F, 53

APSC FILED Time: 8/31/2015 12:11:45 PM: Recvd 8/31/2015 11:16:44 AM: Docket 10-100-R Doc. 133

Protocol G, 62
Protocol H, 72
Protocol J, 74
Protocol K, 89
pump, 36
quasi-experimental, 78
quasi-experimental design, 75
Randomized Controlled Trial (RRCT), 74
RCT, 75
rebate, 17, 18
rebound effect, *see Glossary*
records, 7
regression, 78
regression analysis, *see Glossary*
regression-based, 24
regulatory, 10
reliability, *see Glossary*
reporting, 2, 10
reporting period, *see Glossary*
reports, 83
residential, 16, 33, 37
residential measures, 37
Resource Acquisition Program, *see Glossary*
resources, 7
retrofit, 35, 36
retrofit isolation, *see Glossary*
review, 33
rigor, *see Glossary*
risk, 7, 14
role, 2, 4
SAE, *see glossary*
sample size, 15
sampling, 10, 24
savings, 6, 35
scale, 7
scope, 25
secondary data, 22
Secondary Effects. See Glossary
segmentation, 6
self-report, 22
showerheads, 33
simple engineering model, 22, *see Glossary*
snapback. see glossary
specifications, 35
spillover, 55, *see Glossary*
spot metering, 36
statistical, 24
statistical significance, 81
stipulated values, *see Glossary*
surveys, 7
T-8 lighting, 35
takeback effect, *see Glossary*
technical forum, 69
Technical Reference Manual, 17
time frame, 7
timing, 8, 10, 25, see protocol a
tracking, 8, 17
treatment group, 74, 78, 81
TRM, 17, 33, See Arkansas Technical Reference Manual Vol. 2
types, 2, 21
uncertainty, 14, 35, *see Glossary*
under-performing programs, *see PROTOCOL A*
usage, 7, 33
variable-speed drive, 36
variation in adoption, 79
vendor, 18
verification, 8, 21, 35
water heaters, 33
wattage, 35
weatherization, 16