Northeast Energy Efficiency Partnerships

Decision-Framework for Determining Net Savings Approach

September 2014

Northeast Energy Efficiency Partnerships 91 Hartwell Avenue Lexington, MA 02421 P: 781.860.9177 www.neep.org

REGIONAL EVALUATION,
MEASUREMENT & VERIFICATION FORUM

REGIONAL EVALUATION,
MEASUREMENT & VERIFICATION FORUM

NEEP was founded in 1996 as a non-profit whose mission is to serve the Northeast and Mid-Atlantic to accelerate energy efficiency in the building sector through public policy, program strategies and education. Our vision is that the region will fully embrace energy efficiency as a cornerstone of sustainable energy policy to help achieve a cleaner environment and a more reliable and affordable energy system.

The Regional Evaluation, Measurement and Verification Forum (EM&V Forum or Forum) is a project facilitated by Northeast Energy Efficiency Partnerships, Inc. (NEEP). The Forum's purpose is to provide a framework for the development and use of common and/or consistent protocols to measure, verify, track, and report energy efficiency and other demand resource savings, costs, and emission impacts to support the role and credibility of these resources in current and emerging energy and environmental policies and markets in the Northeast, New York, and the Mid-Atlantic region.

About Johnson Consulting Group

Johnson
CONSULTING GROUP

Dr. Katherine Johnson is President of Johnson Consulting Group, a woman-owned energy efficiency consulting firm based in the Washington D.C. area. For more than 25 years, Dr. Johnson has been providing directing evaluation, measurement and verification (EM&V) studies for electric and natural gas utilities throughout the United States. She is also providing technical support and advice on EM&V best practices for public service commissions in Arkansas, California, Maine, Missouri and Texas.

REGIONAL EVALUATION, MEASUREMENT & VERIFICATION FORUM

Contents

List of Figures

List of Tables

REGIONAL EVALUATION, MEASUREMENT & VERIFICATION FORUM

Decision-Framework for Determining Net Savings Approach

1. Introduction

As energy efficiency (EE) programs continue to increase in importance, measuring the savings directly attributable to those programs has come under increased scrutiny from program administrators, policy makers and key stakeholders.

But it is difficult to measure the unknown — specifically to determine what would have happened in the absence of a program. These effects, known collectively as net savings, include free ridership or attribution, spillover, and market effects, are critical metrics that can define a program's overall success in reaching a specific energy savings goal.

While there are established best practices in the energy efficiency industry for determining net savings, there is hardly consensus on this issue. Rather, measuring the net savings may be approached from several perspectives based on a number of critical variables. The choice of what method or combination of methods and approaches should consider a variety of key issues.
This appendix is drawn on information gathered from a variety of sources and requiring different types of research methodologies and approaches.

This topic has been the subject of several excellent literature reviews and white papers, so for a deeper discussion of these topics please refer directly to the following referenced documents:

- NMR Group, Tetra Tech & KEMA 2011, "*Massachusetts Program Administrators Cross-Cutting Net to Gross Methodology Study for Residential Programs –Suggested Approaches (Final)*, July.
- Violette D. & Rathbun, P. 2014, "*Chapter 17: Estimating Net Energy Savings: Common Practices,"* Uniform Methods Project, September.
- Tetra Tech, KEMA & NMR 2011, "*Massachusetts Program Administrators Cross-Cutting C&I Free-Ridership and Spillover Methodology Study Final Report,* April.

The purpose of this appendix is to increase access to and expand upon the information provided in these reference materials in order to help to inform the decision-making regarding the best approaches to consider when determining program net savings. This document is not intended to be a text book on these measurement techniques for a narrow audience, but rather a framework for a wide range of stakeholders who are involved in making program design and policy decisions.

As a way to facilitate the use of the decision-framework, this discussion begins with a brief description of each of the net savings approach summarized in Table 13 of the Uniform Methods Project, many of which are currently in common use in program evaluations. A copy of this table is provided in Appendix A for reference.

It is important to note that some of these approaches rely on using objective data gathering and analysis, while other techniques are based more on expert judgment. Both analytical techniques have merit. In this discussion, we have grouped these approaches on a continuum identifying those techniques that are driven heavily by data compared to those techniques that are driven heavily by judgment.

The next section highlights the ways these approaches could be used to estimate the three key components of net savings or Net-to-Gross (NTG): Free Ridership; Spillover, and Market Effects. The section summarizes these approaches by both the techniques used as well as by energy efficiency program type.

Section Three provides a discussion of these approaches' strengths and weaknesses to offer additional insight for energy efficiency stakeholders, which include program administrators, and regulators. The discussion of these limitations expands upon the previous literature reviews.

Section Four, summarized from the previous literature reviews, provides additional considerations for evaluators to consider prior to beginning an investigation of net savings.

2. Brief Summary of Net Savings Methods

The following definitions, excerpted from other industry publications, are provided to ensure that all readers have a common understanding of the terms used throughout this appendix.

2.1 Definitions of Key Terms

The factors most often used to calculate net savings are free ridership, spillover (both participant and non-participant), and market effects. These definitions are consistent with those contained in the *Draft Version Energy Efficiency Program Impact Evaluation Guide* and are essentially standard in the evaluation literature (Violette & Rathbun 2014, pp. 4-7 citing (SEE Action 2012).

Free Riders: Program participants who would have implemented a program measure or practice in the absence of the program.

Free Ridership: The percentage of program savings attributable to free riders.

Spillover: Additional reductions in energy consumption and/or demand that are due to program influences beyond those directly associated with program participation. There are generally two types of spillover:

- **Participant spillover:** The additional energy savings that are achieved when a program participant due to the program's influence installs energy efficiency measures or practices *outside* the efficiency program after having participated.
    - *Like spillover:* Refers to program-induced actions participants make outside the program that are of the same type as those made through the program.
    - *Unlike spillover*: Refers to actions participants make outside the program that are unlike program actions but that are influenced in some way by the program.
- **Non-participant spillover:** The additional energy savings that are achieved when a non-participant implements EE measures or practices as a result of the program's influence (for example, through exposure to the program) but is not accounted for in program savings.

Market Effects: Describes a change in the structure of a market or the behavior of participants in a market that is reflective of an increase in the adoption of energy efficiency products, services, or

practices and is causally related to market intervention(s). Some experts suggest that market effects can be best viewed as spillover savings that reflect significant program-induced changes in the structure or functioning of energy efficiency markets (Violette & Rathbun 2014, p. 5).

2.2 Descriptions of Key Methodologies

This section provides a brief overview of the current methodologies used to calculate net savings, including free ridership, spillover, and market effects. It is important to note that not every method will capture all of these effects. This summary is designed to aid the reader in interpreting the summary tables presented in Section 3.

While there are a variety of methods used to calculate net savings, these techniques can be compared across the analytical spectrum. One end of this spectrum focuses on the degree to which the technique relies on objective data compared to the other end which focuses on the degree to which expert judgment or experience determines the outcome. This analytical spectrum is illustrated in Figure 1. The Randomized Control Trials (RCT) is considered to be the "gold standard" for assessing net savings. However, as discussed later in this section, it is not always feasible to use this method. Quasi Experimental Designs (QED) are also data intensive and add a degree of accuracy to these savings estimates.

The Common Practice Baseline, Market Sales Data and Pricing and Elasticity approaches compare data from different jurisdictions to estimate net savings and therefore are more data-centric methodologies.

In contrast, the Structured Expert Judgment and Deemed or Stipulated NTG values rely heavily on judgment-based approaches to provide context to the net savings results. The remaining techniques, including the Counter-Factual Surveys, Top-Down Evaluations, and Historical Tracing (Case Studies) blend a mix of data from program participants with qualitative findings such as those from industry experts. Figure 1 illustrates the spectrum of net savings approaches which are described more fully in the next section.

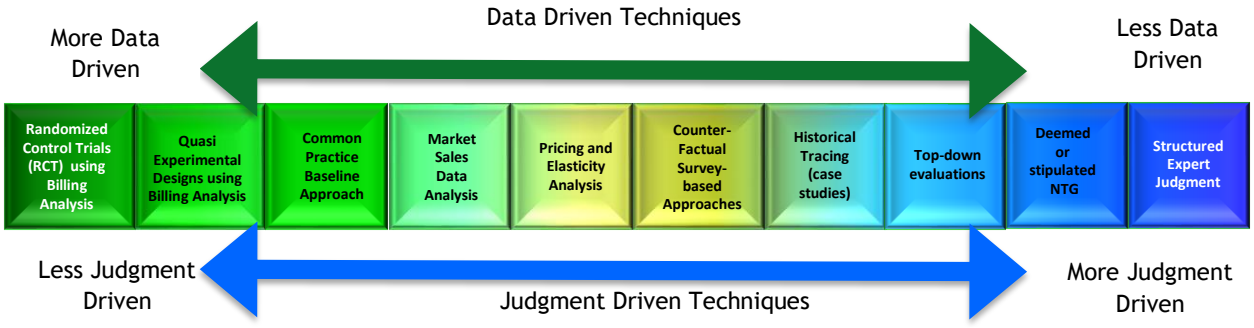

Figure 1: Spectrum of Net Savings Analysis Approaches

Billing Analysis (Regression Analysis): Develops an estimate of program savings by analyzing consumption data. The most common form of such analysis compares usage after program participation with usage prior to program participation, with some form of weather normalization. Since billing analysis typically requires up to 12 months of post-implementation consumption data, this approach may not be feasible where more timely feedback on net savings is required (NMR et al., 2011, p. 4; Violette & Rathbun 2014, p. 59).

- **Randomized Control Trials (RCT):** A study design that randomly assigns participants into an experimental group or a control group. As the study is conducted, the only expected difference between the control and experimental groups in a randomized controlled trial (RCT) is the outcome variable being studied.[1] A billing analysis is commonly used to identify differences between the two groups using a linear fixed-effects regression (LFER) approach, where the regression model identifies the effect of the program by comparing pre- and post-program billing data for the treatment group to the billing data for the control group. A key feature of the LFER approach is the addition of a customer-specific intercept term that captures customer-specific effects on electricity use that do not change over time, including those that are unobservable (Violette & Rathbun, 2014, p. 12). In general, the literature concludes that the RCT approach is viewed as the most accurate method for estimating net impacts (Violette & Rathbun 2014, p. 14).
- **Quasi-experimental Design (QED):** This methodology is used when RCT is not a feasible option. This design involves selecting groups, upon which a variable is tested, without any random pre-selection processes. This form of experimental research used extensively in the social sciences and psychology. Applied to analysis of billing data a quasi-experimental design, consumers typically select themselves into the participant group, and the researcher must then develop the comparison group. To avoid confusion, quasi-experimental designs use the term *comparison group,* and RCT designs use the term *control group* (Violette & Rathbun, 2014, p. 16).

Market Sales Data Analyses (Cross-Sectional Studies): Measures the total net effect of the program, including both free ridership and participant and non-participant "like" spillover. The most common approach is a cross-sectional comparison area method in which post-program data are compared with data from a non-program comparison area (or multiple comparison areas) for the same point in time. Thus, evaluators can make a comparison between the change in the program area from the pre-program period to the post-program period *and* the change in the non-program area over the same period (Violette & Rathbun, 2014, pp. 40-41).

Supplemental Approaches

In addition to the data-driven methods described previously, there are several other approaches that can be used to provide additional guidance in deriving net savings impacts. These approaches rely on collecting and analyzing data from a variety of sources to form an independent estimate of savings. These estimates are based on a mix of data combined with specific judgment or expertise. A summary of these supplemental approaches is provided next.

Common Practice Baseline Approach: These are estimates of what a typical consumer would have done at the time of the project implementation (SEE Action 2012b cited in Violette & Rathbun, p. 34). This approach is still evolving and is being tried in several jurisdictions; however, there is not yet a

[1] http://himmelfarb.gwu.edu/tutorials/studydesign101/rcts.html

consensus on the best strategy for deploying this approach due to both the costs and challenges associated with this technique. The current approach is to estimate common practice baselines through surveys of participants and nonparticipants as well as analysis of market data. But, there is not yet widespread experience in developing common practice baselines allowing for a determination of best practices. So while this approach may be appealing, it is still in development within the EM&V community (Violette & Rathbun 2014, pp. 37-38).

Pricing and Elasticity Analysis: Several methods can be used to determine the effect on purchases of lowering the price through upstream or downstream incentives, also known as price elasticity. These methods are used for programs whose influence on the market is affected chiefly through reducing prices for efficient equipment.

- *Stated Preferences.* Stated preference experiments systematically ask potential customers what they would choose from a set of options with different features and prices, or how a change in price affects purchases. The "choice sets" offered each customer are designed so that the effect of price and features can be estimated from the data set of all the customers' responses.
- *Revealed Preferences.* Revealed preference studies observe the actual choices customers make from true choices available to them when making purchases. (NMR 2011 et al., p. 3).

Survey-Based Approaches (Customer Surveys): Self-Reported counterfactual surveys attempt to determine what would have happened absent the program by asking people what actions they would have taken if the program were not available (for free ridership) and what subsequent actions the program influenced them to take (for spillover) (NMR et al., 2011, pp. 3-4). This approach can be a cost-effective, transparent, and flexible method for estimating NTG, and has become one of the most often used methods in EE net savings estimation. Surveys may focus on three types of respondents: program participants, program non-participants, and market actors (Violette & Rathbun, 2014, p. 22).

Deemed Savings: Deemed or stipulated NTG ratios are typically either set by a regulatory agency or negotiated between regulators and program administrators. These ratios may be determined at the portfolio level or on a measure-by-measure basis. Typically, evaluators base the ratios on NTG studies from past evaluations and/or reviews of other similar programs in which an NTG ratio was estimated. This multiyear estimation of NTG ratios is a compromise between performing net savings estimation studies every year and the use of deemed values based on that research for a selected time period (Violette & Rathbun, 2014, pp. 6, 50)

Historical Tracing (or Case Study) Method: Involves reconstructing the events that led to the outcome of interest using devices typically found in historical studies, journalism, and legal arguments. The program evaluators use historical data or other information from a wide range of sources to develop a "weight of evidence" conclusion regarding the program's influence (Violette & Rathbun 2014, p. 52).
Structured Expert Judgment Approaches (Delphi Panels): Involves assembling a panel of experts who have a good working knowledge of the technology, infrastructure systems, markets, and political environments. This approach is one alternative for addressing market effects in different end-use markets. These experts are asked to estimate baseline market share for a measure or behavior. Structured expert judgment processes use a variety of specific techniques to ensure that the panel of experts specify and take into account key known facts about the program, the technologies supported,

and the development of other influences over time (NMR et al., 2011). The Delphi process is the most widely known technique (NMR et al., 2011 cited in Violette & Rathbun 2014, p. 48).

Top-down Analyses: Use aggregate data that represent the overall level of EE effort across all programs, but cannot isolate the effects of a single program or measure. Top-down models conceptually address all of the NTG factors—free ridership, spillover, and market effects.

2.3 Equations to Estimate Net Savings

Equation 1. Net Savings Including Free Ridership, Spillover, and Market Effects
Net Savings = Gross Savings – FR + SO + ME not already captured by SO Where:
FR = free ridership savings
SO = spillover savings
ME = market effects savings not already captured by SO
Equation 2. Net-to-Gross Ratio
NTG Ratio = 1 – FR ratio + SO ratio + ME ratio (where the denominator in each ratio is the gross savings)

When using the NTG ratio defined by specific free ridership, spillover, and market effect factors (or ratios), evaluators use equation 3 to calculate net savings.
Equation 3. Net Savings Calculation Using the Net-to-Gross Ratio
Net Savings = NTG Ratio * Gross Savings

3. Framework for Estimating Net Savings by Method and Program

Net-to-gross (NTG) effects may be estimated using a variety of different approaches. In deciding which approach or combination of approaches to use in a given situation, several factors need to be considered. Selecting the appropriate net savings analysis methods depends, in part, on the type of questions that need to be answered by a net savings study (Violette & Rathbun 2014, p. 56).

As Figure 1 shows, the level of rigor associated with these methodologies varies based on the type of data collected and analytical methodologies used. The most rigorous methods are based on making assumptions based on data estimates and are described in this appendix as "Estimation Methods." The second type of net savings approaches rely more heavily on direct feedback from either respondents, industry experts, or a combination of the two. These approaches are described as "Assumption Development Processes."

3.1 Criteria for Decision-Making

Selecting the appropriate net savings methodology depends upon a number of variables that play a critical role in determining the suitability or feasibility of each approach. These criteria include the following:

- Primary research objective (i.e., free ridership, spillover, market effects);
- Level of rigor required;
- Cost;
- Availability of key data;
- Availability of suitable comparison groups; and
- Types of programs or measures studied.

The role that these specific criteria play in these decisions is included in the discussion of each methodology, provided next.

3.2 Strengths and Limitations of Net Savings Methodologies

Billing Analysis: Since billing analysis typically requires up to 12 months of post-implementation consumption data, this approach may not be feasible where more timely feedback on net savings is required. In general, billing analysis is appropriate to use only when participant whole-house [or facility] savings are substantial relative to total consumption and when there are large numbers of fairly homogenous participants (NMR et al., 2011, p. 4).

When a comparison group is used in the analysis, the resulting estimate of savings is considered to be net of free ridership; the analysis does not isolate NTG effects from adjustments to gross savings. The comparison group's change in usage is assumed to represent how the participants would have changed absent the program.

This approach depends heavily on the "comparability" of the comparison group. In most programs, customers who choose to participate are different from those who do not participate, in ways that can affect their year-to-year consumption changes. Methods have been developed to adjust for self-selection bias by including in the analysis customer characteristics that might be associated with both the propensity to participate and the consumption changes absent the program, however all such methods rely on identifying assumptions that cannot be empirically validated.

A second situation where a valid comparison is available occurs when customers are randomly assigned to receive the participant "treatment" or not. With random assignment, there is no systematic difference between the untreated or control group customers and the participating customers, other than the treatment itself. Therefore, the control group can provide an unbiased estimate of what participants would have done absent the program treatment (NMR et al., 2011, p. 4), but only if spillover and market effects are assumed not to exist.

Table 1: Summary of Strengths & Limitations for Billing Analysis

Strengths	Limitations	Major Decision Criteria
Is relatively easy to develop a model	Billing analysis savings estimates are frequently limited to aggregate program savings. Regression methods that are designed to provide disaggregated measure savings estimates are subject to modeling error and other sources of bias and can produce unreasonable results (Waite 2013).	Are comparison data available from a homogenous control group? Is the approach designed to minimize self-selection bias? How is the model controlling for error and other types of bias?
Relatively inexpensive compared to other approaches	Billing analysis does not provide any information to account for the realization rate (Waite 2013).	
	Billing analysis is subject to unknown selection and spillover bias (Agnew and Goldberg 2013 cited in Violette & Rathbun 2014).	
	Requires at least one year of data to provide meaningful comparisons	

Randomized Control Trials (RCT): This approach is viewed as the most accurate method for estimating net impacts. The RCT controls for free riders and near-term participant spillover. Although the RCT method can produce an accurate baseline when constructed correctly, it may not be possible to apply an RCT to evaluations of energy efficiency programs for a variety of reasons. RCT generally requires planning in advance of program implementation. Also, an RCT approach may involve denying or delaying participation for a subset of the eligible and willing population. In some cases, the random assignment may result in providing services to consumers who either do not want them or may not use them (Violette & Rathbun, 2013, pp.14-15).

Thus an inherent problem with all comparison group methods (including RCT), distinct from the problem of selection bias, is the inability to discriminate free ridership from non-participant spillover.

Table 2: Summary of Strengths & Limitations for RCT

Strengths	Limitations	Major Decision Criteria
Random assignment reduces and limits bias in estimates	Bias can result if random assignment occurs among volunteers or if program drop out rate differs by key characteristics	Are comparison data available from a homogenous control group? Is the approach designed to minimize self-selection bias? How is the model controlling for error and other types of bias? Are sufficient resources dedicated to ensuring that the model will be properly designed and the experiments properly executed?
Increases reliability and validity	Equity/ethical concerns about assigning some rate payers to a control group and not allowing them to participate in a program for a period of time.	
Controls for free riders and participant spillover	Does not address non-participant spillover and there may result in biased estimates of net savings (Waite 2013)	
Widely accepted in natural/social sciences as "gold standard" of research	Needs to be planned as part of a program implementation to allow for appropriate randomization of participants and control group	
Easier to blind/mask than observational studies	Expensive in terms of time and money	
Results can be analyzed with well-known statistical tools	Volunteer biases: the population that participates may not be representative of the whole	
Populations of participating individuals are clearly identified	Loss to follow-up attributed to treatment	

Quasi-Experimental Designs (QED): QEDs are an improvement over simple comparison approaches. QEDs resemble quantitative and qualitative experiments, but lack random allocation of groups or proper controls, so firm statistical analysis can be very difficult (Violette & Rathbun, 2014, pp. 15, 22). As long as the shortcomings of the quasi-experimental design are recognized, these studies can be a very powerful tool, especially in situations where 'true' experiments are not possible. They are very good way to obtain a general overview and then follow up with a case study or quantitative experiment, to focus on the underlying reasons for the results generated.[2]

[2] https://explorable.com/quasi-experimental-design

Table 3: Summary of Strengths & Limitations for QEDs

Strengths	Limitations	Major Decision Criteria
Limits bias if a matched comparison group can be identified regarding the actions that influence energy use	May be difficult to identify a matched comparison group if there are unobservable variables that affect energy use	Are comparison data available from a homogenous control group? Is the approach designed to minimize self-selection bias? How is the model controlling for error and other types of bias? Are sufficient resources dedicated to ensuring that the model will be properly designed and the experiments properly executed?
Unlike RCT can be applied after program implementation	Does not address non-participant spillover	
Widely accepted in natural/social sciences when random assignment cannot be used	Without proper randomization, statistical tests can be meaningless	
Can be integrated with individual case studies	Results may not stand up to rigorous statistical scrutiny because the researcher also needs to control other factors that may have affected the results.	
	There is no empirical means to determine the adequacy of a comparison group, i.e. there is no way to test the validity of the identifying assumptions (Waite 2013).	
	The assumptions are not sustainable with observable data (Waite 2013).	

3.3 Supplemental Net Savings Approaches

Common Practice Baseline: This approach captures a snapshot of a particular market. But the baseline market will naturally change and evolve over time, so this approach requires making a careful selection of an appropriate market and monitoring its changes over time. This approach is more appropriate for technology driven programs, such as lighting or appliance upstream programs or refrigerator recycling. The most important issues to consider in selecting this approach is to develop a clear understanding of the understand the construction of the baseline used in the evaluation; and to analyze the implications of this baseline against an appropriate counterfactual scenario for that program (Violette & Rathbun 2014, pp. 37-40).

Table 4: Summary of Strengths & Limitations for Common Practice Baseline

Common Practice Baseline Approach		
Strengths	**Limitations**	**Major Decision Criteria**
Can help to avoid double counting of free ridership in circumstances where gross impacts incorporate some net savings factors	Self-selection bias is not addressed and methods for addressing self-selection are not readily apparent	How is the comparison jurisdiction defined? Is the approach designed to minimize self-selection bias? Are the market measures appropriate for upstream analysis (i.e., lighting or appliance programs)? How will this analysis be supplemented with additional approaches to provide context for these findings?
Can be used in upstream programs	Does not capture nonparticipant spillover	
Can be applied market-wide	Common practice baselines for measures and technologies will require updating to reflect market changes This is a relatively new technique that is still evolving and therefore there is no uniform consensus on the best way to execute these studies.	
	This does not capture nonparticipant spillover	

Market Sales Data: The total net effects of a program can be estimated through an analysis of market sales data. The most commonly used approach is a cross-sectional comparison approach in which post-program data are compared with data from a non-program comparison area (or multiple comparison areas) for the same point in time. The NMR et al. (2011) study lists three important factors to consider when determining if this appropriate:

- Does an appropriate comparison area exist?
- Are the market data available and complete?
- What are the features of the program?

Table 5: Summary of Strengths & Limitations for Market Sales Data Analyses

Market Sales Data Analyses		
Strengths	**Limitations**	**Major Decision Criteria**
Can estimate total net effects of a program	There may be low availability and quality of sales and shipment data in the area of interest and appropriate comparison areas	Are comparison data available from a similar jurisdiction? Is the approach designed to minimize self-selection bias? Are the market measures appropriate for upstream analysis (i.e., lighting or appliance programs)? How will this analysis be supplemented with additional approaches to provide context for these findings?
Uses information on actual customer behavior	Data may be expensive to acquire and/or have gaps that are misleading.	
Addresses trends in an entire market	May be difficult to determine the appropriateness of a comparison area, therefore estimated impact is subject to potential bias as with any comparison group approach.	

Pricing and Elasticity Analysis: These approaches rely on customers to identify their intentions, based on actual or stated preferences, compared to hypothetical situations. The major obstacle is that in stated preference models, there is a potential difference between what customers their purchase intentions are in a hypothetical situation compared to what they actually do when given the opportunity (NMR 2011 et al., p. 3). Table 6 summarizes these issues regarding this approach.

Table 6: Summary of Strengths & Limitations with Pricing and Elasticity Analysis

Strengths	**Limitations**	**Major Decision Criteria**
The responses from a group of customers can be extrapolated to the entire population.	There are potential differences between what customers say they will buy in a hypothetical situation and what they actually do buy.	How is the approach accounting for the "halo" or "Hawthorne" effects? Are comparison data available from a similar jurisdiction? Is the program or measure appropriate for this type of analysis (i.e., lighting or appliance upstream programs)? How will this analysis be supplemented with additional approaches to provide context for these findings?
It can measure actual differences in purchase behavior due to changes in price.	It may be difficult to find a valid comparison area for market sales data.	
The researcher can directly observe customers making purchase decisions.	It may require direct observation of sales via onsite visits.	
Suppliers can also provide information regarding the effect of price changes for programs.	May be difficult to differentiate between participants and non-participants.	

(Source: NMR et al., 2011, p. 16)

Survey-Based (Counter Factual) Approaches: Surveys collect NTG-related data. Despite its drawbacks, this approach is typically the most cost-effective, transparent, and flexible method for estimating NTG. Survey based approaches are used in evaluations that start with gross estimates, and then adjust for NTG factors. Surveys may target different types of respondents including program participants, program non-participants, and market actors.

Surveys can yield estimates of free ridership and spillover without the need for a non-participant control group (NMR et al., 2010). However, participant surveys only capture a subset of market effects. Self-reported assessments of program influence are subject to a number of sources of unknown bias that some authors argue generally result in overstatement of free ridership (Peters and McCrae 2008).

Table 7: Summary of Strengths & Limitations Counter Factual Surveys

Strengths	Limitations	Major Decision Criteria
Can provide useful information to support process and impact evaluations.	Potential biases related to "halo" effect	Is the approach designed to minimize self-selection bias? Is the program or measure type suitable for this approach? How will this analysis be supplemented with additional approaches to provide context for these findings?
Flexible approach allows evaluator to tailor question to program design or implementation methods.	Consumers' inability to know what they would have done in a hypothetical alternative situation	
Can yield estimates of free ridership and spillover without the need for a non-participant control group.	The tendency to rationalize past choices	
Is a flexible and transparent method.	Potential arbitrariness of scoring methods based on evaluator judgment	
Is relatively low-cost compared to other methods.	Consumers may fail to recognize the influence the program may have had on other parties who influenced their decisions which in turn impacted the participant.	
	Participant surveys capture only a subset of market effects.	
	There is no way to independently validate the accuracy of the data (Waite 2013).	
	Is best viewed as a supplementary approach in that it is not measuring actual changes in consumption (Neme 2013).	

Structured Expert Judgment Approaches: A particularly useful role for structured expert judging is to develop a "consensus" estimate to consolidate results from multiple estimation methods The Delphi process is the most widely known technique. At least two rounds of judgment are required for a Delphi panel, although more rounds can be used (NMR et al., 2011).

Table 8: Summary of Strengths & Limitations from Structured Expert Judgment Approaches

Strengths	Limitations	Major Decision Criteria
The resulting estimate is the independent, professional judgment of a group of technology and or market experts.	The approach relies on the availability of high-quality data to inform the panel, leading to reasonable estimates of net savings.	Is the approach designed to minimize self-selection bias?
It is a useful approach for programs with diverse and complex end-uses or practices.	Sampling-based calculations of confidence and precision are not available.	Is the program or measure type suitable for this approach?
It is a useful tool for consolidating results from multiple methods to develop a consensus estimate.	Is best viewed as a supplementary approach in that it is not measuring actual changes in consumption (Neme 2013)	How will this analysis be supplemented with additional approaches to provide context for these findings?
Panel members can provide levels of confidence and procedures using appropriate methods.		

Deemed or Stipulated Net-to-Gross Ratios: Deemed or stipulated NTG ratios are predetermined values and do not rely on a calculation-based approach. Deemed values are often based on previous NTG research that was conducted using at least one of the other methods. NTG ratios are often stipulated when the expense of conducting NTG ratio analyses cannot be justified, when the uncertainty of the potential results is too great to warrant a study, or when a preliminary estimate is required pending the completion of primary research.

A recent review of 42 jurisdictions in the United States and Canada (which represented nearly all jurisdictions with ratepayer-funded EE programs) found that only 14% use a deemed approach to NTG for C&I programs compared to 50% of the jurisdictions using an active research approach to developing estimates of net savings factors (Navigant 2013 cited in Violette & Rathbun 2014, pp. 50-51).
Since this method draws from multiple information sources, it is difficult to determine the magnitude of the effects, so the evaluator cannot assign statistical precision to the estimate (NMR et al., 2010) (Violette & Rathbun 2014, p. 53).

Table 9: Summary of Strengths & Limitations for Deemed Savings or Stipulated Approaches

Strengths	Limitations	Major Decision Criteria
This approach can reduce contentious after-implementation adjustments to estimated program savings because agreed upon net savings factors are developed in advance	The process for developing deemed net savings can be contentious.	Is the approach designed to minimize self-selection bias? Is the program or measure type suitable for this approach? How will this analysis be supplemented with additional approaches to provide context for these findings?
The approach leverages information gathered from previous research studies.	It is not based on program-specific information	
	Developing deemed savings values at the measure and technology levels can be expensive and time consuming	
	An incorrect estimate can be deemed.	
	Is best viewed as a supplementary approach in that it is not measuring actual changes in consumption (Neme 2013)	

Historical Tracing (case study method): This method involves the careful reconstruction of events leading to the outcome of interest to develop a 'weight of evidence' conclusion regarding the specific influence or role of the program in question on the outcome. This method is best suited to attribution analysis of major events such as adoption of new building codes or policies, and is not typically applicable to energy efficiency programs (NMR et al., 2011, p. 5).

Table 10: Summary of Strengths & Limitations for Historical Tracing (Case Study Method)

Strengths	Limitations	Major Decision Criteria
This approach can reduce contentious after-implementation adjustments to estimated program savings because agreed upon net savings factors are developed in advance.	The process for developing deemed net savings can be contentious.	Is the approach designed to minimize self-selection bias? Is the program or measure type suitable for this approach? How will this analysis be supplemented with additional approaches to provide context for these findings?
The approach leverages information gathered from previous research studies.	It is not based on program-specific information.	
	Developing deemed savings values at the measure and technology levels can be expensive and time consuming.	
	An incorrect estimate can be deemed.	
	Is best viewed as a supplementary approach in that it is not measuring actual changes in consumption (Neme 2013)	

Top-Down Evaluations (Macroeconomic Models): Top-down methods and program-level evaluation both provide useful, but different, perspectives on the accomplishments of EE efforts (Violette & Rathbun 2014, p. 43)

Top-down evaluations use macro data on energy consumption in a model that relates changes in energy consumption to a measure of energy efficiency program activity, such as expenditures. The top-down approach is appealing since it directly addresses overall net savings. In addition, the regression analyses provide confidence and precision levels around these estimates. This approach, however, is subject to the same bias as any longitudinal regression analysis due to the influence of confounding variables that are not accounted for in the model.

Developing a model that can measure a 1% to 2% change in total energy use annually and is attributable to energy programs requires a reasonably sophisticated structure. Furthermore, the number of observations and quality of data needed to identify a small effect can be challenging (Violette & Rathbun 2014, p. 42).

Top-down approaches seem complementary to results produced by program-level evaluations; however, there may be concerns about using these top-down methods as a replacement for program-level evaluations.

Table 11: Summary of Strengths & Limitations with Top-Down Methods

Strengths	Limitations	Major Decision Criteria
Estimates net effects of all programs cumulatively	Methods are not fully developed at the state or regional levels.	Is the approach designed to minimize self-selection bias? Is the program or measure type suitable for this approach? How will this analysis be supplemented with additional approaches to provide context for these findings?
No need to adjust for free ridership, spillover, or market effects at the aggregate level	Relies on high quality energy consumption data and on data regarding EE efforts within each cross-section analyzed	
Aggregate models can be useful in assessing state and regional environmental impacts such as the impact on carbon emissions.	Cannot provide savings at the measure, technology or program level	
The model can confirm—at an aggregate level—whether the expected energy savings are actually reflected in the macro-consumption data.	Does not provide information on how to improve program design and implementation processes	
	Subject to bias because it assumes that the regression model accounts for all sources of variation in aggregate consumption (Waite 2013)	

3.4 Summary of Methodologies

As a way to illustrate the ways in which these methodological approaches are used to estimate net savings, the appendix provides flow charts illustrating each preferred approach for specific program approaches. These recommendations have been synthesized from the strategies described in the major white papers on this topic (i.e., NMR et al., 2011; TetraTech et al., 2011, and Violette & Rathbun 2014).

The following figures illustrate the alternative approaches to quantifying the net savings for each program type.

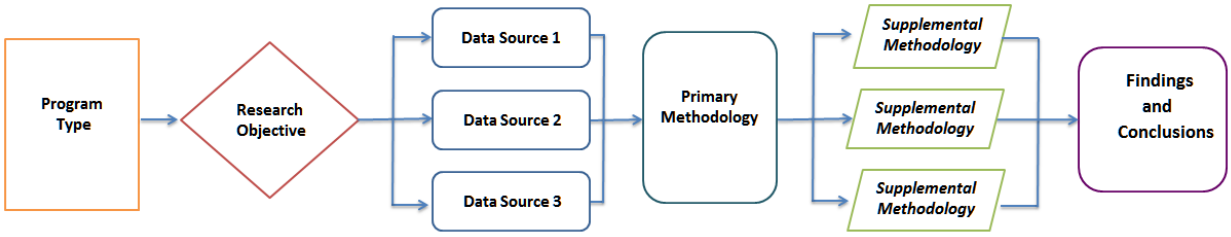

Figure 2: Illustrative Example of Net Savings Methodological Approach

The major challenge with estimating savings from upstream programs, such as residential lighting and appliance programs, is that is difficult to identify program participants. Therefore, rather than being

able to determine program savings from billing analyses or using a RCT or QED approaches or from customer surveys, the only feasible approaches are relying on several methodologies to assess both NTG and market effects. The primary approaches would be to rely on comparisons to sales in similar jurisdictions, accompanied by surveys with retailers or a Delphi panel of experts which can provide NTG estimates based on their expertise. However, it is most important to recognize that determining the NTG, spillover or market effects from this type of program model requires multiple research methods. By applying multiple techniques, the findings provide information to draw conclusions regarding estimates for free ridership, spillover, and market effects (NMR et al., 2011, p. 22).

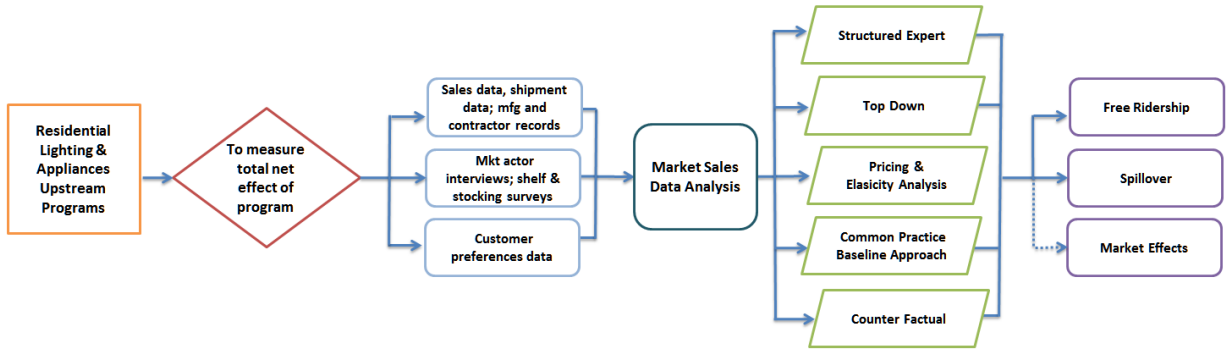

Figure 3: Net Savings Approach for Upstream Residential Lighting and Appliances Programs

To determine net savings from residential programs that provide rebates for specific end-uses such as heating, ventilation, air conditioning (HVAC) or water heating equipment, this approach relies heavily on feedback from customers and contractors. If sales data can be used to estimate free ridership, while market effects can be estimated through interviews with contractors and suppliers in comparison areas. The interviews would gather information on sales levels and market share of efficient and standard equipment. This approach is illustrated in Figure 4.

Using these methods would lead to determining estimates for both free ridership and spillover. Measuring market effects would only be possible if sales data are available to provide valid comparisons.

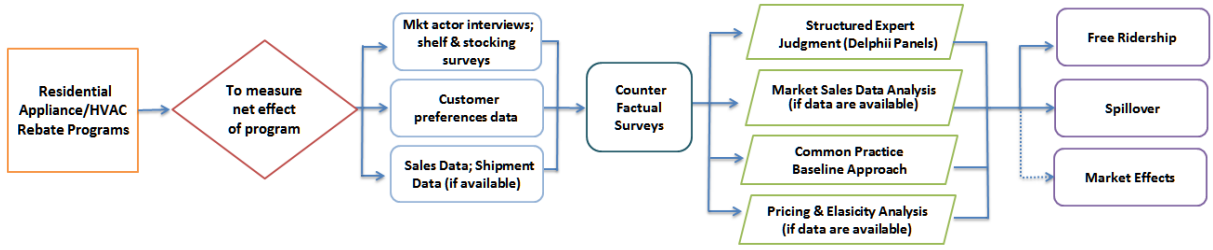

Figure 4: Net Savings Approach for Residential Appliance and HVAC Rebate Programs

To measure the NTG, and spillover from whole house weatherization or retrofit programs, the most appropriate approaches are to rely on customer self-reported counterfactual surveys, supplemented by input by the auditors and contractors, to gauge the effects of the audit and the incentives on

customers' purchase decisions. Using standard billing analysis or stipulated or deemed values may also provide additional ways to validate the findings from these survey-based approaches (see Figure 5).

The result of these net savings approach would provide net savings estimates both free ridership and participant and non-participant spillover. Market effects would be very difficult to quantify for this type of program.

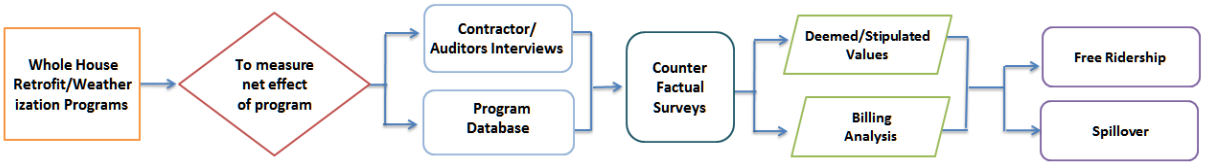

Figure 5: Net Savings Approach for Whole House Retrofit/Weatherization Programs

The whole-house approach used residential new construction programs; also makes sales-based approaches non-viable due to differences in key factors such as building codes and their level of enforcement. Rather, a more appropriate methodology is to rely on surveys with participant builders to estimate free ridership and spillover. To estimate market effects, expert judging via Delphi panels of subject matter experts would be required.

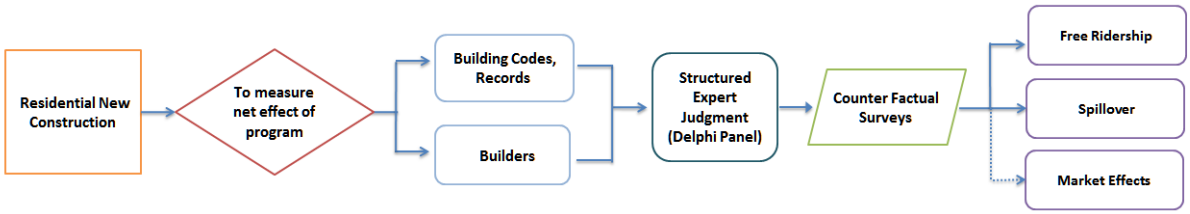

Figure 6: Net Savings Approach for Residential New Construction Programs

Behavioral programs provide additional challenges for measuring market effects since these programs focus on changing participating customers' energy usage by promoting energy-saving behaviors. However, the program design makes the program ideal for a billing analysis approach using RCT which compare the overall energy usage of participants (i.e., those who were randomly selected to receive the Home Energy Reports) with that of the non-participants (a randomly selected group of customers who did not receive the reports) over the same time period. These analyses yield an estimate of program influence on energy savings net of free ridership and spillover.

QED is an appropriate alternative approach to estimate spillover and this information can be supplemented with customer surveys, as Figure 7 illustrates. NMR et al., 2011, pp. 6-8).

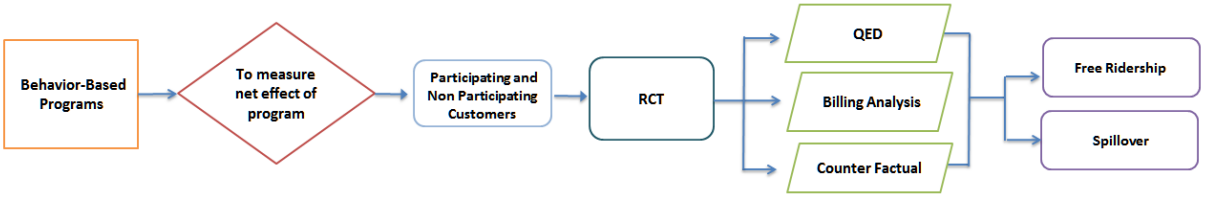

Figure 7: Net Savings Approach for Behavior-Based Programs

The estimation methods are similar for C&I programs as Figures 8 and 9 show. The most feasible approaches to estimate net savings for both direct install and C&I programs are to focus on surveys with decision-makers and market actors (e.g., retail store managers, contractors, etc.). These surveys can be used to gauge the influence of the program on the customer's purchase decisions.

If appropriate comparison groups are available, then billing analysis can be used to provide additional insight into market effects and overall savings from these programs. In some jurisdictions, there may also be deemed savings or stipulated savings for these program types.

However, given the program design is would be difficult, if not impossible, to measure market effects from these types of programs.

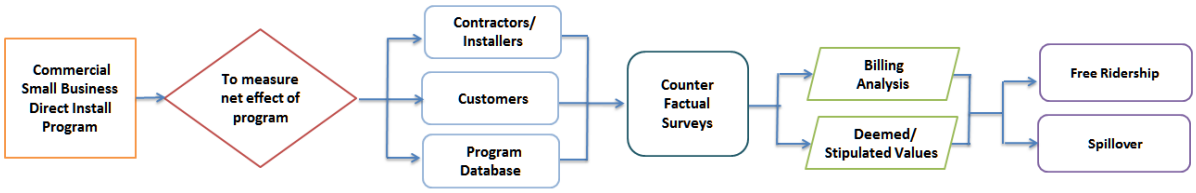

Figure 8: Net Savings Approach for Commercial Small Business Direct Install Program

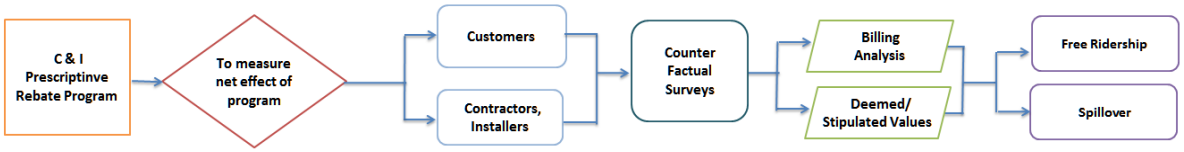

Figure 9: Net Savings Approach for C&I Prescriptive Rebate Program

Custom programs, by their design, usually have few participants. Furthermore, each project will be markedly different from another, so approaches relying on comparison groups are not feasible or practical. Rather, the best approach is to view each project individually, relying on a case study methodology. The free ridership and spillover findings should be further confirmed through supplemental approaches such as a Delphi panel or other type of structured expert judgment. But given the unique nature of this program type, determining market effects is not feasible.

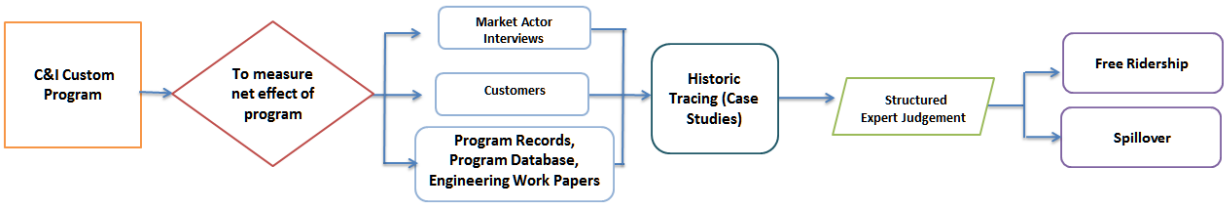

Figure 10: Net Savings Approach for C&I Custom Program

4. Key Issues to Consider Before Beginning a Net Savings Study

Net savings methodologies continue to evolve and improve over time. As the previous discussion illustrates, there is no one single methodology that is appropriate for all programs or measures, and a single methodology is often not the best choice for estimating program or measure net savings. Rather, jurisdictions should design evaluation plans to assess net savings in conjunction with the key stakeholders considering:

- Congruence of the approach with the information required to support a specific policy objective (e.g. resource acquisition, market transformation etc.)
- The appropriate schedule for the evaluation effort over time, taking into account the expected value of the information produced versus the cost of the research effort
- Program design and maturity
- The contribution of the program to overall portfolio savings (past, current, planned)
- The evaluation budget, objectives, and value
- Observations and lessons learned from other jurisdictions.
- Adequately documenting the methods used and communicating the results of any net savings study is important.

It has always been important to consider evaluation options before implementing an energy efficiency program or portfolio of programs. However, planning the types of net savings studies that are needed and the frequency of this measurement prior to program implementation has become critically important. Net savings studies embedded in experimental designs that are established prior to consumers becoming program participants allow for:

- The consideration of randomized designs
- The development of the data platform for estimating consumption-based models (including top-down models)
- The collection of information needed for well-run structured expert panel studies.

Evaluation planners should consider a number of practical issues when planning a net savings evaluation. These include the how the information will be used, program, maturity, study timing, frequency of net savings estimation, and whether to use multiple approaches. The following bullets summarize key issues to consider when selecting a NTG methodology:

- **Use of the information.** It is essential to consider how the results of the net savings evaluation will be used and the audience for which the evaluation is intended.
- **Program maturity.** Almost all programs are assumed to have some free ridership, but free ridership rates will change over the life of the program. For example, as a program matures free ridership will increase, but so will spillover and market effects. Therefore, it is important to understand the life cycle of each program studied to know when it is appropriate to test for test spillover and market effects.
- **Timing of data collection.** To estimate free ridership, the data should be collected as soon as possible after program participation. This minimizes recall bias (Baumgartner, 2013), provides feedback on program design, and reduces the possibility that the key decision-maker or market actor is no longer available. However, if the objective is to estimate spillover, the ideal time to collect data is at least one to two years after program participation, as this allows sufficient

time for spillover to occur. Finally, if the objective is to estimate market effects, then regular data collection over a period of time is required.

- **Frequency of net savings estimation.** The frequency of net savings or NTG analyses depends on the use of the information. If it is a component of financial incentives for a program administrator, evaluators may need to conduct these studies more frequently. Usually, there is no need to perform detailed net savings studies more than every other year.
- **Triangulation of NTG approaches.** Using data from multiple sources may reduce the effects of self-report bias and measurement error (Baumgartner, 2013). Using an in-depth methodology with multiple sources also allows evaluators to weight the value of responses from different decision-makers (Megdal et al., 2009).

Some evaluation issues are best addressed prior to rolling out a new or revised energy efficiency program. Therefore, it is important that program design staff and evaluators work together in advance of implementing a new program to discuss data collection needs for both program implementation and future evaluations.

5. Trends and Recommendations in Estimating Net Savings

As discussed previously, the choice of approach for estimating net savings will vary depending on the questions asked, the characteristics of the program(s) evaluated, and the ultimate use of the data. However, there are trends in the application of methods:

- The expanded use of informational and behavioral EE programs is leading to a greater use of RCTs and quasi-experimental designs that employ some form of randomization to address self-selection.
- The complexity of programs and the need for assessing market effects is leading to a greater use of informed expert panels and Delphi-types of analyses for certain program types as illustrated in Section 4.
- The need to examine trends in program performance over time and impacts on markets over time is resulting in long-term planning for net savings and NTG factor analyses.
- Net savings studies are becoming embedded in survey analyses that are also designed to gather information about program implementation effectiveness.
- The value of information from net savings studies is being considered in a more structured manner to help manage evaluation costs. Both program administrators and regulators often review the appropriate level of confidence and precision to provide fair attribution estimates that minimize risks to both ratepayers and to utilities receiving incentives (Violette & Rathbun 2014, pp. 54-63).

6. References

NMR Group, Tetra Tech & KEMA 2011. "*Massachusetts Program Administrators Cross-Cutting Net to Gross Methodology Study for Residential Programs –Suggested Approaches" (*Final), July. www.ma-eeac.org/Docs/8.1_EMV%20Page/2011/2011%20Residential%20Studies/Residential%20MA%20NTG%20Methods%20Final%20072011.pdf.

Neme, C. 2013. Comments on Estimating Net Savings: Methods and Practices, 4th Draft.

Tetra Tech, Inc., KEMA; &NMR Group, Inc. 2011. *Cross-Cutting (C&I) Free Ridership and Spillover Methodology Study Final Report*. Massachusetts Program Administrators. www.ma-eeac.org/Docs/8.1_EMV%20Page/2011/2011%20Commercial%20&%20Industrial%20Studies/MA%20FR_SO%20CI%20%20Study%20w%20Exec%20Summary%205-26-2011%20v11.pdf.

Violette D. & Rathbun, P. 2014. "Chapter 17: *Estimating Net Energy Savings: Common Practices,"* Uniform Methods Project, September, http://www.nrel.gov/docs/fy14osti/62678.pdf.

The following publications were cited in Violette & Rathbun 2014:

Agnew, K. & Goldberg, M. (2013). "Whole-Building Retrofit with Consumption Data Analysis Evaluation Protocol." Chapter 8 in *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*. www1.eere.energy.gov/wip/pdfs/53827-8.pdf.

Baumgartner, R. (2013). *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures,* Chapter 12: Survey Design and Implementation Cross-Cutting Protocols for Estimating Gross Savings. www.nrel.gov/docs/fy13osti/53827.pdf.

Peters, J. & McRae, M. (2008). "Free-ridership Measurement Is Out of Sync with Program Logic...or, We've Got the Structure Built, but What's Its Foundation." In Proceedings of the 2008 ACEEE Summer Study on Energy Efficiency in Buildings, Washington, DC. www.aceee.org/files/proceedings/2008/data/papers/5_491.pdf.

SEE Action (2012). *Energy Efficiency Program Impact Evaluation Guide*. Prepared by Steven R. Schiller, Schiller Consulting, Inc. www1.eere.energy.gov/seeaction/pdfs/emv_ee_program_impact_guide.pdf.

Waite, S. 2013. Comments on Estimating Net Savings: Methods and Practices, 4th Draft.

Appendix A: Table 12: Uniform Methods Project

Table 12: Summary of Methods Applicable to Different Conditions

Net Savings Method	Surveyed Group	Applicability				Typical Cost or Complexity	Special Requirements
		Custom Measures	Measures With Few, Diverse Participants	Large Numbers of Similar Participants	Measures With Substantial Upstream Influence Invisible to Consumers		
RCTs using DiD	None necessary, but could be conducted to help validate the baseline as an appropriate counterfactual scenario	Poor	Poor	Good	Poor	Low	Random assignment of participants and controls
Quasi-Experimental Design	None necessary but could be conducted to validate or develop better baselines	Poor	Poor	Good	Poor	Low	Matched nonparticipant comparison group
Regression models—Billing data analyses with control variables and Linear Fixed Effects Regression (LFER	Participating consumers and comparison group consumers	Poor	Poor	Good if there is a valid comparison group	Good if there is a valid comparison group	Low	Need control variables that influence energy use across participants and nonparticipants
Survey based—participants, nonparticipants, and market actors	Participating end users	Good	Good	Good	Poor unless combined with retailer or contractor surveys	Medium	Counterfactual baseline based on survey responses
	Participating and nonparticipating end users	Poor	Poor	Good	Poor unless combined with retailer or contractor surveys	Medium-High	Nonparticipants must be representative of participants
	Retail store managers and contractors	Good	Good	Medium	Good	Medium	
Survey based - qualitative sales and counterfactual scenario	Retail store managers and contractors	Poor	Poor	Good	Good	Low	
Structured expert judgment	Experts	Depends on quality of input methods				Low	
Market sales data (cross-sectional studies)	None	Poor	Poor	Good	Good	Low if data are available; high or not possible if data must be developed	Defined market segment
	Manufacturers and regional buyers and distributors	Poor	Poor	Good	Good	Low	
	Retail store managers and contractors	Good	Good	Medium	Good	Medium	
Common practice baseline	Participating and Nonparticipating end-user surveys or market sales data are used	Poor	Poor	Good	Good	Medium to high	Defined market segment
Top-down methods for regional application	None	Requires data on aggregate energy consumption and information on EE effort (expenditures or related program variable) for a large number of cross-sectional observations over a period of time				Depends on the cost of compiling the initial dataset	Aggregate data available on geographic cross-sections

(Source: Violette & Rathbun 2014, p. 59)