Program Administrators of Massachusetts

Net-to-Gross Methodology Research—TXC08

TETRA TECH

NMR Group, Inc.

DNV·GL

March 24, 2017

6410 Enterprise Lane, Suite 300 | Madison, WI 53719
Tel 608.316.3700 | Fax 608.661.5181

tetratech.com

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS/ABBREVIATIONS/DEFINITIONS

Acronyms/Abbreviations	Definition
Net-to-gross (NTG)	The net-to-gross factor indicates how much of the gross savings occurred due to the program and would not otherwise have occurred.
Free ridership (FR)	A program's free-ridership rate is the percentage of program savings attributed to free-riders. A free-rider refers to a program participant who received an incentive or other assistance through an energy efficiency program who would have adopted the same high-efficiency measure[1] on their own at that same time if the program had not been offered.
Spillover (SO)	Spillover refers to additional energy efficiency measures adopted by a customer due to program influences, but without any financial or technical assistance from the program.
Participant Spillover (PSO)	Participant "like" spillover refers to the situation where a customer installed energy efficiency measures through the program, and then installed additional measures of the same type due to program influences. Participant "unlike" spillover is where the customer installs other types of energy efficient measures than those offered through the program, but are influenced by the program to do so.
Nonparticipant Spillover (NPSO)	Nonparticipant spillover refers to any energy efficient measures adopted by program nonparticipants due to the program's influence.
Snapback	Snapback refers to the increase in overall energy consumption after the installation of the efficiency measure, due to increased usage.
Market effects (ME)	Market effects refer to "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)" (Eto et al. 1996)[2].

[1] For purposes of this discussion, an "energy efficiency measure" includes high efficiency equipment or appliances, an efficiency measure such as weatherization, or an energy efficient practice such as turning off a computer when not in use.

[2] Eto, J.; Prahl, R.; Schlegal J. A Scoping Study on Energy-efficiency Market. Transformation by California Utility DSM Programs. Lawrence Berkeley National Laboratory, 1996. http://emp.lbl.gov/sites/all/files/lbnl%20-%2039058.pdf.

1.0 EXECUTIVE SUMMARY

The goals of this study are to examine the general methods used in Massachusetts for estimating net savings and to compare the current methods to methods currently being used in other regions across the country or being considered for use in Massachusetts. We also document the strengths, weaknesses, and integration issues of the various net-to-gross (NTG) methods for different programs and markets in Massachusetts, including a general approach that can be used by evaluators to structure the combining of studies to determine best estimates of NTG. The integration issue is increasingly important given Massachusetts' long history of offering energy efficiency programs, the likely cumulative or market impacts of these energy efficiency programs over time in this region, and the newer evaluation efforts in Massachusetts to account for these cumulative or market impacts.

Two main activities were conducted as part of this study:

1. Review approaches to utilizing self-report information and potential adjustments to current practices in Massachusetts

2. Develop a framework to integrate NTG or net savings results from diverse methods, including self-reports, market effects, and codes and standards evaluation approaches.

The remainder of this executive summary provides additional background and presents the conclusion, recommendations, and considerations for future research. Section 2 summarizes the literature review conducted in other jurisdictions surrounding self-reports, while Section 3 presents considerations and a general approach for integrating NTG and net savings methods. Appendix A provides a summary of NTG methods used in Massachusetts since 2012. This summary, developed during this project, was instrumental in driving the findings and recommendations in this report and may serve as useful reference material for future studies. Appendix B provides examples of integration studies and methods being used in Massachusetts. Appendix C provides a list of the documents cited or referenced throughout the report.

1.1 Background

Research to estimate net savings continues to evolve and adapt to changing program models, market conditions, and external influences. External influences affecting net savings can include consumers' attitudes toward, behaviors regarding, and consumption of energy resources. Over the last several years in Massachusetts, new methods for estimating net savings have been explored and/or used, including the measurement of market effects for sectors where market effects are likely to have occurred or are expected in the future, the measurement of savings from the Code Compliance Support Initiative, and preliminary top-down studies that use an econometric approach to measuring program impacts across a portfolio of programs in a geographic region.

In response, the cross-cutting evaluation team has been investigating the following four different methods or initiatives of estimating net savings over the past three-year cycle in Massachusetts.

- **Self-report NTG studies**, which use a bottom-up approach to estimate net energy savings for individual measures/end-uses, programs, or groups of programs. Specifically, these studies provide estimates of free-ridership (FR) and like spillover (LSO) at the end use, program, and PA level through interviews with program

participants and trade allies. Nonparticipant spillover (NPSO) is also captured through interviews with participating or non-participating vendors who are asked about sales of program-qualifying equipment that was sold but not through a PA program. The self-report method is among the most widely used methodologies for determining net energy savings for downstream programs across the nation.

- **Market effects studies** rely on an analysis of program-qualifying equipment sales compared to all equipment sales, as well as self-report data on free ridership and spillover. These studies measure net SO savings from program-induced changes in the structure and functioning of the market, as well as seeking evidence of such changes. Market effects measurement recognizes that programs can drive nonparticipant savings through market effects that are not captured by program project measure tracking or participant NTG evaluation. Market effects, if identified, may justify claiming additional program savings.

    - **Code Compliance Support Initiative (CCSI).** The code compliance evaluation seeks to measure net savings attributable to the CCSI for improving code compliance in Massachusetts through various avenues over the long term, and to gather supporting evidence for those savings. Evaluation activities include implementing immediate surveys after each classroom training and reporting the findings at the end of the contract period, implementing follow-up interviews with training attendees, examining building code compliance documents, estimating the portions of Massachusetts code officials who have attended different types of trainings; and planning for attribution likely to take place in the late 2017 and early 2018. Codes and standards create a confounding influence on top-down approaches, self-report NTG, and market effects, as savings may also be driven by compliance with codes and standards that are due to the PAs' efforts in improving compliance.

- **Top-down modeling** techniques use an econometric approach to estimate program impacts across all energy-efficiency programs in a given geographical region or service territory rather than running separate studies for each program (or measure/end use within a program). Top-down approaches use regression models attempt to measure changes in energy consumption over time that are attributable to programmatic interventions by the PAs. In theory, top-down methods are capable of capturing the full portfolio-level effect, including FR, SO, market effects, and snapback across multiple programs. Top-down modeling is currently in the research and development stage in Massachusetts and it is not yet clear whether this method will be viable for estimating net savings.

These four cross-cutting topic areas are interrelated. In theory, the top-down evaluation area describes the net effect of all the PA programs and efforts on changes in total energy consumption. However, the underlying drivers of these net savings are better understood by using other methods to estimate net savings. These include participant NTG surveys, estimation of market effects, and estimation of savings attributable to codes and standards support. Taken together, these four topic areas identify program-driven savings in programs and markets in areas that overlap and provide different explanations for the net savings.

For this study, the evaluation team worked with Program Administrators (PAs) and Energy Efficiency Advisory Council (EEAC) members in a collaborative forum to identify and prioritize the myriad issues and approaches related to net savings techniques. The collaboration followed two parallel paths. First, self-report approaches were researched to identify recent industry techniques or changes that may be useful to consider for Massachusetts. Second, the general issue of approaches, issues, or practices of integrating NTG or net savings studies was

researched to provide explanation and guidance to PAs and evaluators seeking to combine or reconcile the findings from diverse techniques. The self-report approach research has immediate opportunities for adopting or testing new techniques, survey questions, or algorithms in Massachusetts. The integration research has longer term application for leveraging the best self-report results and merging them with market effects or other studies that capture different information across a broad scope of evaluation studies.

The results presented in this report provide a background on past NTG evaluation history in Massachusetts and forward-looking guidance on the opportunities to evolve approaches. The findings and recommendations presented below provide for near-term opportunities the PAs can act on. These are followed by additional considerations to frame future activities.

The broader discussion and guidance found in the next two detailed sections of this report and appendices provide a larger set of general perspectives and information that can be used to shape future evaluation and research. Taken as a whole, the opportunities presented in this report identify ways to better understand the full impact of programs on participants, markets, and the utility system as a whole.

1.2 Conclusions and Recommendations

In general, participant free ridership (FR) and short-term like participant spillover (PSO) have a well-established history and practice in Massachusetts. However, the commercial and industrial (C&I) evaluation framework (established in 2003 and updated in 2011) and the residential framework (established in 2011), do not include the full range of evaluation methodologies currently in use in Massachusetts. NTG and net savings methods such as pricing and elasticity analysis, approaches to market effects, and top-down econometric modeling are not included in the 2011 NTG framework documents[3] for either the residential or C&I sectors.

As discussed in this document and the appendices, there are many possible combinations of study methods and results that evaluation teams may want to integrate to arrive at a final NTG or net savings result. The challenge for integration comes when attempting to quantify nonparticipant spillover (NPSO) or larger market or sector effects using additional methods, where the scope of analysis of the additional market studies do not necessarily align with those of the FR/PSO studies. The NTG Collaborative has identified a repeatable and transparent general approach to how evaluation teams can structure the integration of multiple NTG and net savings studies into results that combine the qualities of each study, leveraging the strongest results from each while identifying points of weakness.

The general approach organizes results from each study in a table[4] that shows findings for each net savings component as well as the qualities or key considerations of each study's results. The details captured are as follows:

- The result of specific NTG components captured in a study or how cross-component results combine to inform the specific component. For example, is SO differentiated between participants and nonparticipants, or was only one type of SO researched

[3] Tetra Tech; KEMA; NMR Group, Inc. (2011). Cross-Cutting (C&I) Free-Ridership and Spillover Methodology Study Final Report. Massachusetts Program Administrators, and NMR Group, Inc. and Tetra Tech (2011). Cross-Cutting Net to Gross Methodology Study for Residential Programs – Suggested Approaches. Final report prepared for the Massachusetts Program Administrators.

[4] For more details, see Section 3.2 and Table 1-**Error! Main Document Only.**. Example of Hypothetical Summary Information Used for Integration.

- Overall NTG or net savings results
- Statistical confidence and precision, if available
- Any concerns an evaluator may have about over- or under-stated results, including judgements made with qualitative or subjective information
- Baseline differences between studies.

Evaluators would use the summary information to inform what parts of each study might be used to inform final NTG or net savings results. The tables might be shared within an expert panel to ensure consistent information is being considered by each participant in the panel and may also be used to identify critical gaps in the information that suggest a need for additional specific research.

> **Recommendation 1:** Either consider this document to be a supplement to the 2011 framework documents or preferably, update the 2011 NTG framework documents to reflect the current range of practices adopted or being explored in Massachusetts for self-report methods, and include the general approach to the integration of multiple methods presented in this report. Due to the complexity of integration issues and likely variability in the types of studies being integrated, Massachusetts should avoid being overly restrictive or prescriptive in its application of the general approach to integration.

The C&I Baseline Working Group is establishing a C&I baseline framework that will set industry standard practice (ISP) baselines for specific measures and define when non-ISP baselines should be used. The guidance in the C&I baseline framework applies to both electricity- and natural gas-saving measures, both prescriptive and custom measures, and all C&I Massachusetts programs.

> **Recommendation 2:** Based on the recommendations in the final baseline framework, evaluators will need to update question wording surrounding the efficiency portion of the current NTG questions asked of customers and vendors. The C&I Baseline Working Group also agrees that for true custom measures that do not have a recognizable market and thus needs a project-specific baseline, both NTG and the baseline work should be conducted at the same time, on the same sample by the same contractor team. Additionally, algorithms used to develop NTG estimates should be updated to align with market actor responses to the ISP or “unique” baselines.

1.3 Additional Considerations

In review of the different methodologies employed across the nation, the PAs could consider analyzing existing data files to determine if changes to the existing C&I self-report approach (SRA) survey documents are needed. These considerations include:

- **Testing a shorter instrument.** The current C&I battery of questions is long, and for those customers who install multiple measures through the programs, or install measures at many facilities, the questionnaire can be even more burdensome. While a shorter set of questions will reduce respondent burden, this may result in less comparability of results over time. A sensitivity analysis could be done on prior years’ C&I survey data to examine how using a subset of questions and/or excluding consistency checks would impact the results.
- **Revisiting how missing data is handled.** The cross-cutting evaluation team could revisit the handling of missing data (e.g., don’t know responses) to understand the

impact of the adjustment. Some jurisdictions, including Massachusetts use a mean substitution approach to impute missing data, while others prefer different methods (Pennsylvania TRM recommends a linear regression). Prior data files could be re-analyzed to understand the extent and impact of different imputation methods.

- **Review the impact of trade ally responses.** Prior data files of customer and trade ally responses could be analyzed to revisit how trade ally responses are integrated with customer reports and their impact on NTG. For example, an analysis could look at the impact of not replacing participant FR scores with vendor responses in cases where customers report the vendors as influential. Analysis could also look at the impact of potentially capping FR scores in situations where vendors are not answering for a customer-specific project.

The PAs could also consider separating the measurement of FR and PSO to minimize recall bias and allow for sufficient time for participant spillover to happen by:

- **Surveying participants for FR soon after installation and later for PSO.** In the most recent Massachusetts C&I study, two waves of FR data collection were conducted in order to survey customers closer to their participation to minimize recall bias. In addition, a separate PSO study was conducted, allowing customers more time after program participation to realize PSO due to program effects. If the PAs continue to implement the NTG studies using this approach, which is a best practice, we recommend modifying the frameworks to specify the timing of the FR measurement closer to customer participation (e.g., at a minimum within six months of participation) and conducting a separate participant PSO study at least one year after participation.
- **Providing sampling guidance in situations where there is a small number of program participants and/or a small number of participants installing certain measures.** These strategies could include prioritizing FR or PSO based on program characteristics and goals, using deemed or stipulated values for one of the components, collecting data for each component in alternating years, or using non-SRA methods for assessing one of the components.

Other considerations include:

- **Expanding the residential framework to include specific questions and algorithms for different program types (e.g., gas, downstream, audit-based).** Unlike the C&I framework, the residential framework does not specify the SRA questions and algorithms for measuring free-ridership (FR), participant spillover (PSO), and nonparticipant spillover (NPSO). If this were to be done, any updates should be prioritized based on programs or measures with the highest contribution to savings. As a recent Iowa report[5] observes, standardizing the SRA questions and algorithms would help to "ensure that differences over time or between program administrator service territories are due to actual differences in program attribution as opposed to differences in research methodologies." While the Massachusetts PAs apply the same approach for a given program across the state, approaches could vary across similar program types so providing guidance on questions and algorithms would ensure consistency.

[5] State of Iowa, Department of Commerce. Iowa Energy-Efficiency Net-to-Gross Report, September 2, 2015.

- **Provide guidance on how to apply net savings or NTG results from market effects studies**. Because market effects studies focus on the market rather than the program, this approach may capture more savings than the participant focused approach, thus providing a more complete picture of savings influenced by the program. Therefore, the PAs should consider how to apply net savings or NTG results at a sector or portfolio level.

2.0 SELF-REPORT APPROACH METHODS

As stated in the Net-to-gross (NTG) Methodology Research Stage 3 Plan,[6] one of the goals of that study is to review recent NTG self-report approach (SRA) methods used in other regions across the country and compare them to those used in Massachusetts. This section is focused on the SRA and the questions and algorithms used for the different documents included in our review.

2.1 Background

In 2003 and 2011 for C&I programs, and in 2011 for residential programs, the cross-cutting evaluation team conducted net savings methodology studies by reviewing methodologies in use across the nation at that time. These reviews explored the pros and cons of alternative methods for estimating what would have happened absent the program in different contexts. Further, the reviews provided recommendations for conducting net savings studies in Massachusetts (MA) with a focus on SRA methods. These recommendations are referred to as the Massachusetts NTG frameworks.

The methodology for the previous SRA NTG studies of C&I downstream programs in Massachusetts followed a standardized methodology first established in 2003 and updated in 2011[7] for use in situations in which end users were able to report on program attribution via self-report methods. This approach to estimating free-ridership (FR) and spillover (SO) consisted of sequential questions used to identify FR and SO. Respondents were asked about their decision-making process and then asked about the actions they would have taken if the program services had not been offered. This approach also assessed the program's impact on project timing, measure quantity, and efficiency levels while explicitly recognizing that the cost of energy-efficient equipment can be a barrier to installation in the absence of program administrator (PA)-sponsored energy efficiency programs. These questions were followed by questions about additional equipment purchased since participating in the program. Contractor, manufacturer, and vendor feedback and sales data were also utilized to enhance end-user feedback for a number of C&I studies, including the High Bay Lighting (HBL) Market Effect Study (2011).[8]

The SRA approach was also recommended in the Massachusetts residential sector,[9] although the framework did not specify a standard algorithm or set of survey questions. In addition to the SRA, other net savings methodologies were recommended in the residential framework and have been used. For example, the net impact evaluation of the Home Energy Ratings program used self-reports and a billing data analysis to arrive at net savings.[10] The Home Energy Services (HES) NTG evaluation used participant and nonparticipant self-reports as well as

---

[6] Net-to-Gross Methodology Research Stage 3 Plan—TXC 8, February 19, 2016. Prepared for the Massachusetts Program Administrators.

[7] Tetra Tech; KEMA; NMR Group, Inc. (2011). Cross-Cutting (C&I) Free Ridership and Spillover Methodology Study Final Report. Massachusetts Program Administrators.

[8] HBL Market Effects Study. June 17, 2011. http://ma-eeac.org/wordpress/wp-content/uploads/High-Bay-Lighting-Market-Effects-Study-Final-Report.pdf.

[9] Tetra Tech, NMR Group, Inc., KEMA (2011). Cross-Cutting Net to Gross Methodology Study for Residential Programs – Suggested Approaches. Final report prepared for the Massachusetts Program Administrators.

[10] Navigant and Illume (2015). *Massachusetts Cross-Cutting Behavioral Program Evaluation OPower Results.* Final report prepared for the Massachusetts Program Administrators.

discrete choice modeling to estimate the impacts of the HES program's marketing and incentives on consumer decisions.[11] A net savings study of the Residential New Construction program relied on structured expert judgment.[12] For residential lighting, the residential framework suggested the use of multiple methods to estimate savings followed by the use of structured expert judgment to reconcile the results and develop consensus estimates. This was due to several factors unique to the design of the upstream residential lighting program, such as the large amount of savings involved, difficulties inherent in the evaluation of upstream programs, rapid market change, and the consequent risk of being wrong in either direction. This multi-method approach was followed in a recent residential lighting study,[13] including the additional use of structured expert judgment to estimate prospective net savings.

Alternative SRA approaches and methods have gained increasing attention in the industry in the years since these frameworks were last developed. These include, but are not limited to: (1) employing periodic (monthly, bimonthly, or quarterly) telephone surveys to capture NTG feedback, intended to maximize recall and decision maker availability, (2) improving data gathering techniques to support discrete choice analysis, and (3) exploring alternative survey batteries and scoring algorithms. This current methods review focuses on the SRA techniques being utilized across the nation.

2.2 Research Objectives

This study builds on the recent Massachusetts SRA approaches to identify the current state of NTG methods, identify opportunities for using new or different methods, assess strengths and weaknesses, and understand how best to make use of the estimates that have been or will be developed.

Our review focused on the NTG methodologies recommended in seven states over the past five years. These states were selected due to their recent efforts to standardize methodologies utilizing the SRA approach. The methodologies reviewed include:

- Illinois Technical Reference Manual (TRM)
- Fast Feedback Program Rollout (funded by the Energy Trust of Oregon)
- Evaluation Framework for Pennsylvania
- Arkansas TRM
- Iowa Energy-Efficiency Net-to-Gross Report Recommendations
- New York Evaluation Guidance
- Wisconsin Focus on Energy Net Savings Methodologies.

For each study, we summarize the recommended approaches and the key differences from the Massachusetts SRA framework. We also summarize any information provided in the

[11] Cadmus and Navigant (2012). *Home Energy Services Net-to-Gross Evaluation.* Final report prepared for the Massachusetts Program Administrators.
[12] NMR Group, Inc. (2014). *Residential New Construction Net Impacts Report.* Final report prepared for the Massachusetts Program Administrators.
[13] NMR Group, Inc., Cadmus, and DNV GL (2015). *Multistage Lighting Net-to-Gross Assessment: Overall Report.* Final report prepared for the Massachusetts Program Administrators.

approaches on recommended timing of data collection as well as any discussion on baseline issues.

2.3 Illinois TRM

Illinois Statewide Technical Reference Manual for Energy Efficiency, Version 5.0. Volume 4: Cross-Cutting Measures and Attachments. February 11, 2016.
http://www.ilsag.info/il_trm_version_5.html

The Illinois (IL) Commerce Commission directed the evaluation teams to compile and formalize consistent NTG methods for use in IL EM&V work. The Commission's directives were twofold. (1) assess NTG methodologies and survey instruments that have been used to evaluate energy efficiency programs, and (2) compile the most justifiable and well-vetted methodologies in an attachment to the updated IL TRM. The Commission noted that the IL NTG Methods should be flexible and adaptable to multiple program designs and budgets. It also noted the Methods should be tailored to appropriately assess the specifics of each of the Program Administrators' energy efficiency programs. The resulting statewide NTG methodology document covers the majority of residential and nonresidential programs offered in IL. If the NTG protocol is no longer appropriate, instructions are included for diverging from the IL NTG Methods.

2.3.1 Overview of Residential NTG Approaches—Illinois TRM

The IL TRM includes a residential cross-cutting NTG protocol as well as protocols for specific residential programs, including Appliance Recycling, Upstream Lighting, Prescriptive Rebate, Single Family Home Energy Audit, and Residential New Construction. The cross-cutting residential protocol formulates the core NTG as 1 – free-ridership (FR) + participant spillover (PSO) and provides specific questions and scoring algorithms for measuring FR and PSO. It also provides specific questions and algorithms for measuring nonparticipant spillover (NPSO) from trade allies and customers, implying that they are to be included in the core NTG formula. *This cross-cutting protocol provides detail on measuring PSO and NPSO, but it defers to the specific program protocols for measuring FR.* Unlike the Massachusetts residential framework, the IL TRM provides specific questions and algorithms for measuring PSO and NPSO.

The specific protocols for programs include:

- The Appliance Recycling protocol includes basic and enhanced SRA methods with specific questions and scoring algorithms for measuring FR including questions on how the appliance would have been disposed of in the absence of the program. The enhanced method may include additional research methods such as a retailer survey, appliance market assessment survey, or nonparticipant survey. The protocol does not provide specific guidance for when to use each SRA method nor for measuring SO, and thus the cross-cutting protocols may be assumed to prevail.
    - *Noteworthy Differences from Massachusetts.* The IL TRM provides an Appliance Recycling Protocol, which is not included in the Massachusetts framework.
- The Residential Upstream Lighting protocol recommends using store intercept surveys for the customer SRA to measure PSO and NPSO. The protocol includes specific questions and scoring algorithms for measuring these NTG components. It includes specific questions for measuring FR and allows for partial FR. These include questions to assess program influence (captures the maximum level of program influence, reported by a survey respondent, of the residential lighting program on their decisions to

purchase program bulbs on the day of the survey) and no-program questions (used to estimate how many program bulbs a survey respondent would have purchased in the absence of the residential lighting program); FR is calculated as the average of the responses to the two questions.

    - *Noteworthy Differences from Massachusetts.* The IL Residential Upstream Lighting protocol specifies a single data source, store intercept surveys, whereas the Massachusetts framework recommends the use of multiple methods followed by a Delphi panel consensus approach.

- The Prescriptive Rebate with No Audit protocol provides basic and enhanced methods with specific questions and scoring algorithms for measuring FR. Questions include program influence and no-program components[14] as well as consistency check questions on the program's influence to resolve possible conflicting responses. The basic method measures FR using a customer SRA. The enhanced method provides a protocol to triangulate and develop a weighted combination of FR estimates from two sources: the basic method and a trade ally survey. When multiple methods are used, evaluators may triangulate results by rating the analysis methodology and data collected using responses (rated on a scale of 0 to 10) to three questions: how likely is the approach to provide a more accurate estimate of FR, how valid is the data collected and the analysis performed, and how representative is the sample. The weight for each method is the average score for that method divided by the sum of the scores for all methods.

    - *Noteworthy Differences from Massachusetts.* The IL Prescriptive Rebate (with no audit) protocol specifies questions and algorithms for measuring FR—a basic method based on customer SRA and an enhanced method based on trade ally SRA. This is generally similar to the approach in Massachusetts for the Residential Heating and Cooling and the Heating and Hot Water Equipment programs. However, it is different from the recommended approach for the Massachusetts ENERGY STAR® Appliances Program evaluation, which uses multiple methods to derive several NTG estimates followed by a Delphi panel consensus approach.

- The Single-Family Home Energy Audit protocol provides specific questions and scoring algorithms for measuring FR with different approaches for free/direct install versus rebated/discounted measures. The protocol measures FR using a customer SRA with questions on installation timing, quantity, and no-program scenario. Program influence questions are excluded for free/direct install and included for rebated/discounted measures. It also includes consistency check questions on the program's influence for rebated/discounted measures to resolve possible conflicting responses.

    - *Noteworthy Differences from Massachusetts.* The IL Single-Family Home Energy Audit Protocol provides specific questions and algorithms for measuring FR and SO from participants, nonparticipants, and trade allies. The approach in Massachusetts for the MassSave and Weatherization Program evaluation is slightly more complex and recommends using customer self-reported

[14] Respondents are asked to report their likelihood (using a 0 to 10 scale where 0 is "not at all likely" and 10 is "extremely likely") to implement specified energy efficiency measures in the absence of the program. That likelihood score is then divided by 10 to produce the no-program score.

counterfactual surveys, supplemented by input by auditors and contractors to gauge effects of audits and incentives on customers' purchase decisions.

- The Residential New Construction protocol recommends using builder surveys for the participant SRA to measure FR, PSO, and NPSO. The protocol includes specific questions and scoring algorithms for measuring these NTG components. The protocol measures FR using a participant SRA with questions on program influence installation timing, quantity, and no-program scenario. It also includes consistency check questions on the program's influence to resolve possible conflicting responses. PSO includes additional questions to help estimate amount of savings using IL TRM protocols, such as quantity of appliances or location and amount of insulation. NPSO is based on surveys of two groups: dropout builders not participating in the past 12 months and true nonparticipating builders.
    - *Noteworthy Differences from Massachusetts.* The IL TRM approach to measuring NTG for Residential New Construction specifies questions and algorithms for measuring FR, PSO, and NPSO. The approach in the Massachusetts framework does not specify NPSO measurement; however, the Massachusetts framework recommends a more complex approach overall based on: (1) self-report counterfactual surveys with participant builders that can be used to estimate FR, (2) expert judging (Delphi panel) to assess attribution of observed changes in building practices (from baseline studies), followed by (3) building modeling to estimate counterfactual energy usage based on Delphi panelists' estimates.

2.3.2 Overview of Nonresidential NTG Approaches—Illinois TRM

The IL TRM includes a core NTG protocol for nonresidential programs as well as protocols for specific programs, including the C&I New Construction, Small Business, and Study-based programs (e.g., programs that include an energy audit or assessment). There are core protocols for FR, PSO, and NPSO that provide specific questions and scoring algorithms associated with calculating FR and SO scores. That said, the core net-to-gross ratio (NTGR) for an energy efficiency program is defined as 1 – FR even though they define PSO and NPSO.

The core FR protocol comprises three scores: Program Components FR Score, Program Influence FR Score, and No-Program FR Score, each ranging from 0 (no FR) to 1 (full FR). The three scores are combined to calculate the FR value. They are calculated as follows:

Program Components FR Score: Participants are asked to rate the importance of various factors on the decision to implement energy efficiency measures. The numeric scales range from 0 to 10, where 0 means "not at all important" and 10 means "extremely important." The factors included in the survey are program and non-program factors that could impact the participant decision-making process. The evaluator can calculate the score in one of two ways:

1. Equal to 1 - ([Maximum Program Factor Rating]/10).
2. Equal to 1 - ([Maximum Program Factor Rating]/([Maximum Program Factor Rating]+[Maximum Non-Program Factor Rating])).

Program Influence FR Score: Respondents are asked to allocate 100 points to the program and to non-program factors. The points the participants allocate to the program are the "Program Points." The "Program Influence FR Score" is calculated as 1 - (Program Points/100).

No-Program FR Score: Respondents are asked to report their likelihood (using a 0 to 10 scale where 0 is "not at all likely" and 10 is "extremely likely") to implement specified energy efficiency

measures in the absence of the program. That likelihood score is then divided by 10 to produce this score.

The TRM states that consistency checks should be included in the survey questions to check the consistency of the FR responses. The protocol also provides guidance around vendor influence, including when and how to incorporate vendor responses into the FR calculation. The TRM outlines three scenarios to help decide when to utilize vendor responses, which is based on how involved the trade allies are in the program (i.e., integral in the delivery; part of a select, pre-approved network; implement projects and submit applications on behalf of the customer; sign agreements with the program administrator; or complete program-sponsored training). If vendor surveys are used, the TRM outlines questions that can be asked and based on the responses, when the results would be incorporated. Based on three scenarios, the evaluator decides if the vendor rating should be considered a program factor or non-program factor.[15]

- The Small Business protocol follows the core nonresidential FR protocol but includes a few exceptions primarily to reduce respondent burden.
    - *Noteworthy Differences from Massachusetts.* In IL, to reduce respondent burden, the survey can be shortened to remove the program influence FR score from the algorithm, as influence of non-program factors are included in the program components FR score. In MA, small businesses follow the same NTG battery, although the timing score is adjusted due to the shorter planning schedule.
- The C&I New Construction protocol follows the core nonresidential FR protocol but removes the timing aspect, as the program typically does not impact the acceleration of the construction.
    - *Noteworthy Differences from Massachusetts.* In IL, the timing element is removed. In MA, new construction follows the same core question wording and algorithm as other programs.
- The Study-based protocol follows the core nonresidential FR protocol but includes additional questions about maintenance and performance of the measure.
    - *Noteworthy Differences from Massachusetts.* The IL FR algorithm is an average of three independent program elements (program components, program influence, and no-program score), which differs from MA's handling of the timing, quantity, and efficiency elements where one score is calculated.

2.3.3 Timing

The residential cross-cutting protocol states that FR questions should be asked near the beginning of the participant survey, before satisfaction questions. It also states that when estimating SO based on trade ally surveys, respondents should be allowed sufficient time to collect data to inform their responses and not rely on guesses.

The nonresidential core protocol does not provide direction on the timing of the FR survey. However, for SO, the protocol states the PSO module can be implemented as part of the NTG survey or separately, but timed to allow sufficient time—a minimum of three months—after program participation to allow for SO to occur.

[15] Illinois Statewide Technical Reference Manual for Energy Efficiency, Version 5.0. Volume 4: Cross Cutting Measures and Attachments. February 11, 2016, Page 34.

2.3.4 Baselines

In a general discussion of NTG methods, Common Practice Baseline Approaches are discussed but they are not included in the recommended approaches for IL. In a discussion on measuring Residential and Nonresidential New Construction NPSO, the TRM notes that evaluators should make efforts to ensure the building code under enforcement for each jurisdiction is used as the baseline when evaluating SO savings, which differs from the C&I Baseline Framework that uses ISP.

2.4 Energy Trust of Oregon

Fast Feedback Program Rollout: Nonresidential and Residential Program Portfolio; Research Into Action (funded by Energy Trust of Oregon), December 31, 2010.

The Energy Trust of Oregon implemented a methodology to utilize Fast Feedback Methods for collecting feedback shortly after program participation. Monthly samples from 23 participant groups are drawn to collect program satisfaction as well as purchasing decision questions to produce estimates of FR. (Customers are excluded if they had been contacted for an Energy Trust survey in the previous 12 months.) Estimates of FR were based on: (1) how the project would have changed without program assistance, (2) the availability of funds to do the project without program assistance, and (3) the program's influence on the project. These elements were then used to create a project change score and program influence score to produce a FR estimate.

The survey lists possible participant actions, and asks the participant to confirm what actions he or she would have taken if the Energy Trust incentive had not been available. For all groups except Refrigeration Recycling, the options included: the program-supported activity would not have taken place at all; the activity would have been postponed more than one year; and the activity would have been undertaken in exactly the same way as it had been done with program support. These options indicate that lack of program support would have resulted in maximum change (the first two options) or no change (the third option) compared to what was done with program support. Responses indicating that a lack of program assistance would have resulted in significant changes to what was done and high program influence on the project were taken to indicate a lack of FR. Responses that indicated that a lack of program assistance would not have changed what was done and that there was little program influence on the project were taken to indicate the presence of FR. The response options for the refrigeration recycling program were different from all other programs and indicated that a lack of program support would have resulted in either a complete change from what they did with program support or no change at all.

Respondents were also asked to rate the influence of several program elements on how the project was done. Influence was rated on a five-point scale, from "1" (not at all influential) to "5" (extremely influential). In addition, the survey for residential and nonresidential participants in the Solar Electric program asked about the influence of participation in a community-driven solar effort, and the survey for participants in the Refrigeration Recycling program asked about the influence of the free pickup and removal of the recycled refrigerator. For the nonresidential New Buildings program, response options were modified so that an option of "would not have done commissioning" was included. The response option of "purchased less expensive equipment" was removed, as this does not apply to new buildings, and replaced with "reduced energy design features." Additional questions and question modification options are provided for specific programs and equipment. The report also provides guidance on producing estimates in situations where participants were not able to provide feedback on each element.

For each respondent, two scores were calculated—a Project Change Score and a Program Influence Score. Both scores ranged from “0” (indicating no FR) to “50” (indicating high FR).

- The algorithm for calculating the Project Change score was specific to the survey group, but used the following general logic. If the project would have been cancelled or changed significantly without program influence, the score is 0, indicating no FR. If the project would have changed somewhat, but retained some energy efficiency features results, the score is 25, indicating moderate FR. If the project would have changed little or not at all, the score is 50, indicating high FR. For nonresidential customers, the Project Change Score was based on responses to the project change question and the availability of funds question.
- The Program Influence Score was based on the highest rated influence from among the various program elements rated. This score has five possible values: a rating of “5” (highest possible influence rating) results in a score of 0, indicating no FR; a rating of “4” results in a score of 12.5, indicating low FR; a rating of “3” results in a score of 25, indicating moderate FR; a rating of “2” results in a score of 37.5, indicating high FR; a rating of “1” (lowest possible influence rating) results in a score of 50, indicating complete FR.

For each individual, the Project Change and Program Influence scores were summed. The resulting summed score ranged in value from 0 to 100 and was interpreted as a percentage indicating overall FR. If an individual did not provide sufficient data to calculate either a Project Change Score or a Program Influence Score, two overall FR scores were estimated: 1) a low-scenario score, which assumed that the missing score (Project Change or Program Influence, whichever could not be calculated) was 0; and 2) the high-scenario score, which assumed that the missing score was 50.

Spillover and vendor surveys are not included as part of the fast feedback methods.

- Noteworthy Differences from Massachusetts. The Fast Feedback approach consists of a streamlined series of questions so the battery is shorter than the Massachusetts survey and contains no consistency checks. The Oregon nonresidential survey asks customers if their firm had funds available to cover the entire cost of the project and the question is utilized in the algorithm. While a similar question is asked in MA, the question is not part of the FR scoring algorithm. The Fast Track approach does not include SO or vendor surveys and is focused on FR estimation, while the Massachusetts guidelines take into account SO and the use of vendor information when they have been involved in the decision-making process.

2.4.1 Timing

To improve the reliability of satisfaction and FR data, the Energy Trust recommended surveys on a rolling basis throughout the year, collecting participant feedback shortly after completion of program-assisted projects. Monthly samples are drawn and survey results are tracked across the programs.

2.4.2 Baselines

No baseline discussion.

2.5 Evaluation Framework For Pennsylvania

Evaluation Framework for Pennsylvania Act 129 Phase III Energy Efficiency and Conservation Programs, May 31, 2016.

As noted in the Evaluation Framework for Pennsylvania:

"When conducting NTG research, the NTG methods should be consistent across time and EDCs [electric distribution companies].[16] If the NTG metric is measured the same way every year or every quarter, program staff can use the NTG metric to inform their thinking because it provides a consistent metric over time. Another reason for a uniform NTG approach is that the value that can be obtained from comparing NTG metrics across utilities. Just as programs change year to year, it is clear that the programs offered by the EDCs vary from each other. When there are different metrics, no one can discern whether different NTG values are due to program differences, external differences, or differences in the metric. By using a consistent metric, program staff can at least rule out differences in the metric as the reason."[17]

This rationale is consistent with that recommended in the Massachusetts Frameworks. The Evaluation Framework for Pennsylvania further states:

"The SWE[18] has determined that EDCs should use survey methods for assessing free-ridership and spillover for downstream programs and has provided descriptions of common methods for doing those assessments (Appendix B, Appendix C, and Appendix D); these approaches must be used for the specific programs they apply to, though they may be used in combination with other methods. The SWE has established a procedure whereby EDCs may identify downstream programs for which the common methods are not suitable; in such cases, EDCs may propose a method, subject to SWE review. In Phase III, the EDCs may use methods of their own choice, including market effects approaches, to estimate NTG for upstream programs. The EDCs will be provided a common set of methods for upstream lighting programs. The common upstream lighting program NTG methods allow some flexibility for individual EDCs. They include Market Progress Indicators (MPIs) to assess overall market progress, and options for NTG methods."

2.5.1 Overview of Residential and Nonresidential NTG Approaches—Pennsylvania Evaluation Framework

In Pennsylvania, the Evaluation Framework follows Pennsylvania Act 129 in requiring NTG measurement for the purposes of program planning but not for compliance with targets. The Evaluation Framework for Pennsylvania includes a *common approach* that is similar to the one developed by the Energy Trust of Oregon. This approach is designed to minimize respondent burden, and uses a short battery of questions with no consistency checks. The Evaluation Framework for Pennsylvania delineates variants to the common approach for direct install and financing programs and provides specific NTG protocols to measure FR for Appliance Recycling and for Upstream Lighting. The Evaluation Framework for Pennsylvania also provides guidance for prioritizing NTG measurement for High Impact Measures. In addition to the use of SRA in the common approach, the Evaluation Framework for Pennsylvania also discusses the use of econometric approaches and market effects studies. Because of the cost of econometric approaches and market effects studies and the inability to disaggregate the effects for FR and

[16] Electric Distribution Companies (EDCs).
[17] Evaluation Framework for Pennsylvania Act 129 Phase III Energy Efficiency and Conservation Programs, May 31, 2016.
[18] Statewide Evaluator (SWE).

SO, the SWE has determined that survey methods should primarily be used for measuring FR and SO for downstream programs, and that other methods can be used for upstream programs or to provide information on market effects.

The common approach formulates the core NTG as 1 - FR + SO. It provides a SRA to measure FR for downstream programs, specifying survey questions to assess intention (no-program) and program influence; additional guidance is provided on customizing the approach, e.g., assistance programs or programs with multiple market actors as well as guidance for measure-specific questions. The guidance also specifies SO questions for both PSO and NPSO along with residential and nonresidential wording differences. Although NPSO questions are described for both customer and trade ally self-reports, estimation of NPSO is described as desirable but not required. The Evaluation Framework for Pennsylvania also prescribes scoring algorithms for each of these NTG components. The FR algorithm allows for identifying partial FR.

The guidance explicitly specifies the exclusion of consistency checks for assessing FR using a SRA. Citing the Sage Publication's *Handbook of Theories of Social Psychology,*[19] it states that the "assumption that the inconsistency can be resolved accurately may be unfounded. That assumption is based on the belief that the questioner can accurately and reliably determine which of two inconsistent responses is the correct one. A respondent confronted with inconsistent responses may seek to resolve the consistency, but that does not mean that the final response will be accurate. Instead, the response may be influenced by "self-enhancement" motivation." The Evaluation Framework for Pennsylvania goes on to observe that "Other reasons not to confront respondents with inconsistent responses are that doing so may make respondents feel uncomfortable, and as a result, it could color later responses; it also lengthens the survey. Lengthening the survey, and perhaps even inducing some discomfort, may be acceptable if the result is better data. However, as argued above, there is reason to believe that it will not do so. Further, the need to assess which response is correct brings more evaluator subjectivity into the assessment. Therefore, we recommend against consistency checks." [20]

2.5.2 Residential Approaches

Unlike in Massachusetts for the residential programs, the Evaluation Framework for Pennsylvania provides specific questions and algorithms for measuring PSO and NPSO and also specifies conditions and approaches for combining trade ally and PSO to estimate NPSO. Specifically, it states that "If responses to the trade ally survey indicate that the trade-ally-identified commonly sold program-rebated measures comprise a large percentage (e.g., 90 percent or more) of all high-efficiency equipment sold, then evaluators should attempt to determine what percentage of the total trade ally-identified SO is from nonparticipants by subtracting the total participant SO for that sector from the total trade ally-reported SO."[21] In addition, although the Massachusetts Framework mentions market effects in multiple places, it does not specifically discuss market effects studies.

[19] Swann, William B., Jr. "Self-Verification Theory." In P. Van Lange, A.W. Kruglanski, and E.T. Higgins (eds.), *Handbook of Theories of Social Psychology*. Thousand Oaks, CA: Sage Publications, 2011.
[20] Evaluation Framework for Pennsylvania Act 129 Phase III Energy Efficiency and Conservation Programs, May 31, 2016, Appendix C, Section C.6.4, page C-16.
[21] *Ibid,* page D-11.

- For Direct Install programs, in addition to the common approach, the Evaluation Framework for Pennsylvania outlines additional questions to reflect the unique elements of such a program (e.g., influence of any contractors and the provision of measures at no cost.)
- The Pennsylvania Appliance Recycling protocol includes specific questions and scoring algorithms for measuring FR; it does not provide specific guidance for measuring SO. The Evaluation Framework for Pennsylvania also cites the Uniform Methods Project (UMP) and says that a secondary source of supporting data may come from nonparticipant surveys. To maintain consistency with UMP, it recommends averaging results of nonparticipant and participant surveys. Use of a nonparticipant survey is recommended but not required given budget and time considerations.
    - *Noteworthy Differences from Massachusetts.* The Evaluation Framework for Pennsylvania provides an Appliance Recycling Protocol, which is not included in the Massachusetts residential framework.
- The Pennsylvania Upstream Lighting protocol is very similar to the approach used in Massachusetts and recommends the use of multiple methods followed by a Delphi panel consensus approach.
    - *Noteworthy Differences from Massachusetts.* The Residential Upstream Lighting protocol emphasizes the use of a common set of metrics or MPIs to feed directly into NTG estimates or indirectly to help determine NTG—and in particular, market effects—by providing insight into such metrics as efficient lighting availability and awareness. The Massachusetts Framework suggests using multiple methods to derive several NTG estimates. These estimates would be presented to a Delphi panel of experts, who would come to a consensus on the overall NTG estimate for the program. Estimation methods could include: (1) market data analysis on purchase/sales data collected from customer self-reports of purchases and interviews with vendors and suppliers, (2) in-store revealed preferences observations, and (3) shelf and stocking surveys of retail stores, paired with price elasticity analysis.

2.5.3 Nonresidential Approaches

Similar to Massachusetts, the Pennsylvania TRM provides specific questions and algorithms for measuring FR, PSO and NPSO and also specifies conditions and approaches for combining trade ally and PSO to estimate total SO. Question wording and response options should be tailored to the program or measure type such as for Direct Install programs, where the Pennsylvania TRM outlines additional questions to reflect the unique elements of such a program (e.g., influence of any contractors and the provision of measures at no cost.)

- Noteworthy Differences from Massachusetts. The Pennsylvania surveys consist of streamlined series of questions so the battery is shorter than the Massachusetts survey. The Evaluation Framework for Pennsylvania indicates imputing of means for “don’t know” responses is not preferred (as it inflates the FR estimate for cases where the mean FR is less than 50 percent) and instead recommends linear regression to predict intention score from the influence score (mean imputation is used in MA). When asking about a customer’s past participation there is mention that regulators limit consideration to the current year or phase and that it may be difficult to determine whether program influence was from that current year or from an earlier year.

2.5.4 Timing

The Pennsylvania common approach does not provide any guidance related to the timing of measurement.

2.5.5 Baselines

In a discussion of Market Effects studies, the Pennsylvania common approach notes the importance of measure-specific or program-specific baseline measurements for comparison and that they should be broad enough to cover possible interactions with other external influences. For the purposes of measuring NTG, baseline measurements are noted to be the counterfactual.

2.6 Arkansas TRM

Arkansas Technical Reference Manual, Version 6.0, August 31, 2016. http://www.apscservices.info/EEInfo/TRM6.pdf

The Arkansas TRM protocols are used to determine NTG ratios by isolating FR and SO. For all customer classes except self-directing customers[22], the protocol lists five approaches for determining NTG. However, the question sets and final methods used to conduct the NTG analysis are provided by the EM&V contractor.

2.6.1 Overview of Residential and Nonresidential NTG Approaches—Arkansas TRM

The Arkansas TRM NTG 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).[23]" The protocol provides a decision tree for determining when and how to conduct NTG research (see Figure 2-1 below). In general, it formulates the core NTG as 1 – FR + SO and mentions five measurement approaches: (1) self-report surveys of participants and nonparticipants; (2) enhancements to the self-report surveys through review and analysis of interview, documentation, and market-based sales data; (3) econometric methods; (4) deemed; and (5) stipulated NTG ratios. However, the approach recommended by the Arkansas TRM includes analysis of utility consumption data and the enhanced self-report method using other data sources relevant to the decision to install/adopt. The protocol neither requires use of trade ally responses nor specifies algorithms to integrate them, if obtained, but mentions that any use of trade ally responses should be discussed within the evaluation plan.

[22] Eligible nonresidential customers can "opt-out" of utility provided energy efficiency programs, and instead participate in a self-directed energy efficiency option.

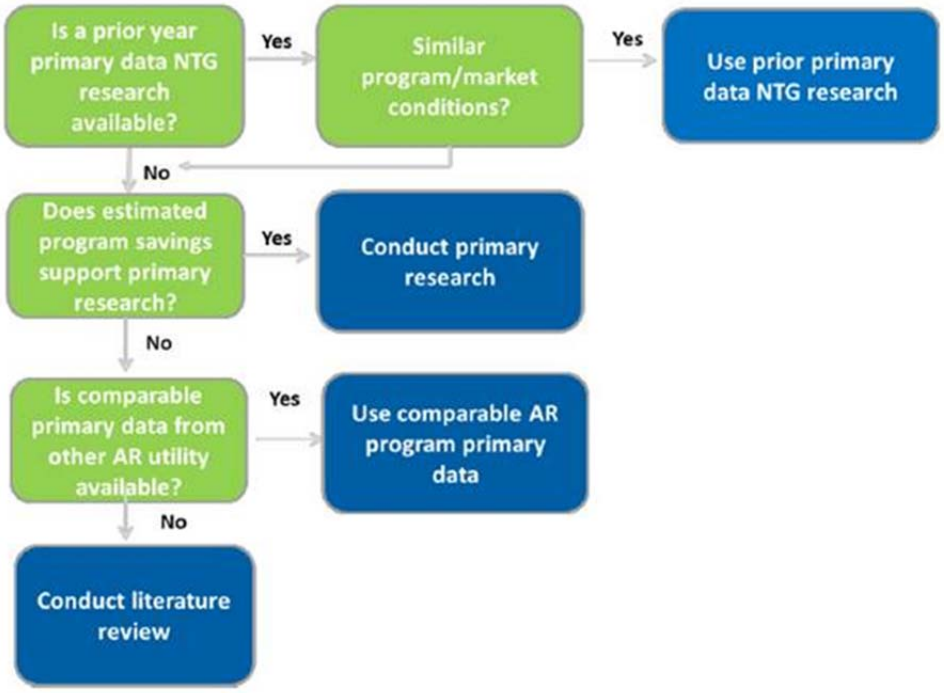
[23] *Arkansas Technical Reference Manual, Version 6.0,* August 31, 2016, Section II – Protocol F, page 49. LCFC is the reduction in potential utility sales revenue resulting from decreased sales due to EE measures implemented under utility funded EE programs.

- The Arkansas TRM specifies FR survey questions to assess intention (no-program) and program influence; responses to influence questions are analyzed for each respondent to identify whether the direct responses on FR are consistent with how each respondent rated the "influence" of the program. It also specifies a three-step approach to measuring PSO, which includes determining whether SO exists at all, the extent of the SO, and the amount of savings per SO project. Although more relevant for C&I programs, the protocol identifies participant internal and external SO. It also mentions NPSO but does not specify applicable questions. The Arkansas TRM does not specify algorithms for measuring SO or FR and categorically excludes identifying partial FR.
    - *Noteworthy Differences from Massachusetts.* The Arkansas protocol provides a general recommendation of triangulation of multiple data sources including on-site inspections and surveys from other jurisdictions. Guidance is provided, although evaluators can deviate within reason from the guideline and document reasons for this within their evaluation plans. The Massachusetts guidelines recommend triangulation as well, albeit in the form of surveys with vendors involved in the decision and expert panels and for specific programs such as lighting and residential new construction.

2.6.2 Timing

As specified in the Massachusetts Frameworks, the Arkansas TRM notes that FR measurement is most accurate when collected as closely as possible to the time of participation decision, whereas SO measurement is most accurate when collected after sufficient time has elapsed that SO activities can have been undertaken. In AR, no specific timeframes are mentioned; however, unlike in MA, the Arkansas TRM provides a decision tree for determining when and how to conduct NTG research (see below).

Figure 2-1. Decision Tree for the Timing and Selection of NTG Research from the Arkansas TRM[24]



2.6.3 Baselines

The Arkansas TRM does not discuss baselines in the context of NTG measurement.

2.7 Iowa Energy-efficiency Net-to-Gross Report

State of Iowa, Department of Commerce. Iowa Energy-Efficiency Net-to-Gross Report, September 2, 2015.

The three investor-owned utilities (Alliant Energy, Black Hills Energy, and MidAmerican Energy) have been using a deemed NTG of 1.0 for all energy efficiency programs for more than a decade. In 2015, the Iowa Utilities Board (IUB) approved a collaborative process to provide more complete and accurate information on NTG. This report documents NTG estimation methodologies currently being used across the country and provides context and guidance on which methods are most appropriate to the Iowa programs and provides guidance and recommendations for Iowa stakeholders to consider in developing Iowa-specific NTG approaches.

2.7.1 Overview of Residential and Nonresidential NTG Approaches—Iowa

The Iowa NTG report identifies three perspectives stakeholders can have on net savings, attribution, and research:

- Energy resource view. This perspective tends to emphasize estimating net savings. The focus is on programs as resource investments similar to traditional supply-side

[24] Ibid, page 55.

resources. Program impacts are measurable in terms of both gross and net savings. The use of retrospective NTG application would be most consistent with the resource view perspective—savings should be real, quantifiable, and only claim what is actually occurring.

- Market view. This perspective emphasizes examining the overall market while still recognizing the importance of achieving savings from energy efficiency programs and activities. NTG research should focus on understanding the markets in which energy efficiency and other demand-side programs operate, particularly the program's influence on the market. Common practice market baseline values and NTG values may be generated through market-based research. The goal is to understand energy efficiency markets as comprehensively as possible in relation to programs, and use this information to provide program design guidance. The use of prospective NTG application would be more reflective of the market view perspective (i.e., NTG has some uncertainty, and what is most important is impacting the market).
- Deemed view. "A deemed perspective suggests that there is enough information in the market to estimate net savings ex-ante by looking at current sales data, and from studies in other regions such that net savings is estimated at a level of accuracy acceptable to the stakeholders. This viewpoint may be driven by perceptions that NTG studies are expensive to conduct, are subject to biases, and do not add enough additional accuracy to the ex-ante deemed estimates of net savings. The use of prospective NTG application would be more reflective also of the deemed view (i.e., NTG has significant uncertainty so it's essential to lock down values prior to program implementation)."[25]

The report states that "because most portfolios offer such a diversity of programs, the recommended NTG methods have generally selected certain program types (e.g., downstream rebates) in prescribing the more detailed NTG approaches"[26] and that the "primary reason for developing common approaches to NTG estimation is to help ensure that differences over time or between program administrator service territories are due to actual differences in program attribution as opposed to differences in research methodologies." It also notes that the "primary drawback of common approaches, however, is that they can be perceived as inflexible, and thus inappropriate for programs that do not exactly fit a typical model. They can also be seen as stifling innovation for new and potentially superior approaches."

The Iowa NTG report found that "nearly two-thirds (62 percent) of jurisdictions that use net savings allow for free-ridership, participant SO, and nonparticipant SO, while 21 percent allow for free-ridership and participant SO but do not allow for nonparticipant SO. Only 17 percent of the jurisdictions with net savings (a total of four states) limit net savings to net of free-ridership (i.e., do not allow for contributions from spillover to count toward the net savings estimates)."[27]

The report does not discuss specific question batteries to measure NTG components. Regarding algorithms, however, it notes that "another means of increasing confidence in the results—or at least build consensus on the efficacy of the approach—derives from explicitly identifying assumptions relating to NTG calculations. Algorithms include relatively arbitrary adjustments that are based on negotiated assumptions around how survey answers should be

[25] State of Iowa, Department of Commerce. Iowa Energy-Efficiency Net-to-Gross Report, September 2, 2015, page 22.
[26] Ibid, page 44.
[27] Ibid, page 43.

treated rather than empirical data. Sensitivity analyses are often performed to assess the influence certain assumptions have on the final values of a free-ridership algorithm."[28]

Triangulation of methods is also mentioned in the report as current industry best practice and that it is "increasing in popularity due to its ability to enhance the rigor of results, and to better represent instances of overlapping energy resource and market NTG viewpoints."[29] The report goes on to mention this practice is used to limit bias and measurement error because it allows evaluators the ability to weight estimates from different methods differently based on the perceived likelihood of bias and reliability of the estimates produced.

The report recommends that for some minor programs, such as education, tree-planting, low-income and some multifamily programs, a deemed NTG of 1.0 is appropriate and no formal NTG research is necessary.

- Noteworthy Differences from Massachusetts. In Iowa, when looking at the timing (acceleration) there are no differences for small business, whereas in MA, the planning cycle for small businesses is taken into account. The timing questions have ranges of months, while the Massachusetts survey asks respondents for a specific number of months/years and uses that in the algorithm. The Iowa example algorithm includes a question about when customers learned about the rebate (e.g., before or after installation) and this question is used in the example algorithm. This question is not asked in the Massachusetts C&I and residential surveys.

2.7.2 Timing

In Iowa, it is recommended that NTG research be conducted once per each five-year planning cycle. Programs contributing large savings to the portfolio or programs in rapidly changing markets may need primary research to be conducted every two to three years and possibly more frequently. Research findings will provide direction on when this additional or new NTG research should be done. There is no mention of the length of time from participation to measuring NTG.

[28] Ibid, page 32.
[29] Ibid, page 31.

2.7.3 Baseline

Describing Common Practice Baseline approaches, the Iowa NTG report states that such approaches "determine gross savings relative to what customers would have commonly done in the absence of the program, implying that it is also an ex-ante estimate of net savings and that little (if any) further adjustments are necessary to obtain ex-post net savings. Common baseline approaches are used in the Northwest, but are not common elsewhere. Still, a key insight from these approaches is that the definitions of baselines as defined in TRMs can affect the way gross and net savings are estimated. The premise behind this method is that a "current" baseline can be estimated ex-ante, i.e., before the program is implemented. These baselines are most often determined by the equipment that is typically purchased in the market at the current time, and defined as the average energy efficiency of currently purchased equipment. This is similar to what is outlined in the C&I Baseline Framework. A variant is to use the value of the highest efficient unit available, or at least a unit with higher-than-average efficiency."[30]

2.8 New York Evaluation Guidance

New York Evaluation Plan Guidance for EEPS Program Administrators, August 2013.

Prior to the introduction of the NY Reforming the Energy Vision (REV) in 2015[31] (changes associated with the REV are noted below), the NY provided evaluation plan guidance for program administrators. "The Evaluation Plan Guidance for EEPS Program Administrators (Guidance Document) required that evaluations estimate both gross and net energy and demand impacts. Various methods exist for estimating net energy and demand impacts, including true experimental design, quasi-experimental designs and the self-report approach (SRA) among others. The first two approaches estimate net energy and demand impacts directly. The SRA approach is used to estimate a NTGR, an index of program influence, and defined as 1- FR + SO."[32]

While the SRA is the method most frequently used and is the method the guidelines are focused on, the approach is not mandated, allowing evaluators to propose a method they consider most appropriate during the evaluation planning:

> *"Most evaluation plans and completed reports that have been reviewed by the New York Department of Public Service have relied on the SRA method. The SRA is a mixed methods approach that uses, to varying degrees, both quantitative and qualitative data and analysis to assess causality. However, in these reviews, DPS[33] has discovered that in both the residential and nonresidential sectors, the SRA method is not always designed and implemented according to best practices. Thus, the DPS has chosen to develop the Guidelines for Estimating Net-to-Gross Ratios Using the Self-Report Approach (SRA Guidelines) that requires analysts to address certain key issues but does not require analysts to address these issues in a specific way. The primary use of these SRA Guidelines is to assess the influence of the program on measures installed through the program and to make sure that evaluators are adhering, whenever possible, to these best practices. The Guidance Document does not mention all the available methods and leaves it up to the evaluators to select the method that is most appropriate*

[30] Ibid, page 30.
[31] http://www3.dps.ny.gov/W/PSCWeb.nsf/All/CC4F2EFA3A23551585257DEA007DCFE2?OpenDocument.
[32] New York Evaluation Plan Guidance for EEPS Program Administrators, August 2013, Appendix G, page 1.
[33] Department of Public Service (DPS).

and expects them to adhere to best practices. Finally, the Guidance Document does not preclude the estimation and inclusion of broader PA-induced market effects in place of SO."[34]

2.8.1 Overview of Residential and Nonresidential NTG Approaches—New York Prior to REV

The guidelines provide direction on sections and questions to consider when developing the questionnaire including identifying the correct respondent, set-up questions to get the respondent thinking about the events that led to the installation, and decision-making questions to help them answer the timing, quantity, and efficiency questions for a specific measure. The guidelines recommend using multiple questions to measure FR rather than relying on a single survey question. In addition, the guidelines include discussion around including consistency checks and their use and considerations for missing data. Guidance is also provided for producing participant and nonparticipant estimates to ensure that there is no double counting.

- *Noteworthy Differences from Massachusetts.* The types of questions to include are similar in both states. However, the Massachusetts C&I Framework is more prescriptive in the wording of these questions.

2.8.2 Timing

Prior to the NY REV, the New York guidelines indicated that the number and frequency of evaluations conducted for a program will depend on the type and size of the program, the evaluation cost, and other factors, such as the level of rigor of previous evaluations. The evaluation timetable should be included in any evaluation plans.

To minimize recall bias, the guidelines recommended that FR interviews be conducted soon after their participation in the program. To assess SO, it is recognized that interviews will be conducted later to allow enough time for customers to install additional equipment.

2.8.3 Baseline

No baseline discussion in the context of NTG measurement.

2.8.4 Recent Changes

Con Edison has changed their approach to estimating NTG due to the introduction of REV. Prior to NY REV, the evaluators used a structured battery and algorithm approved by the Commission. Since the inception of NY REV, as an example, Con Edison is using a more streamlined set of questions (five to six questions) for each program and incorporating the NTG questions into their rolling quarterly participating customer satisfaction surveys. The battery does not check for consistency and does not estimate SO.

- *Noteworthy Differences from Massachusetts.* The Con Edison surveys consist of streamlined series of questions so the battery is shorter than the Massachusetts survey and contains no consistency checks or SO questions.

[34] New York Evaluation Plan Guidance for EEPS Program Administrators, August 2013, Appendix G, page 1.

2.9 Public Service Commission of Wisconsin—Focus on Energy

Focus on Energy Calendar Year 2014 Evaluation Report Appendices, May 27, 2015, Appendix L: Net Savings Analysis Methodologies

The Focus on Energy (Focus) report documents two different approaches that were used to assess net savings for the 2014 evaluation year for residential and nonresidential programs—standard market practice and SRA. Using the SRA, SO is combined with the program-level FR results to achieve the NTG ratio, using the following calculation: NTG = 1 – FR + SO.

2.9.1 Overview of Residential and Nonresidential NTG Approaches—Wisconsin

Using the standard market practice (SMP) methodology, data was collected "through the evaluation process to define the average market baseline and average program-installed energy consumption (kWh and/or therm) of each measure category."[35] The methodology goes on to say "The SMP methodology calculates net savings as the difference between the average market baseline and the average program-installed energy consumption, under the assumption that FR is captured in the baseline. To calculate the NTG ratio using the SMP method, the evaluation team compared the net savings determined through the SMP analysis to the verified gross savings."

In 2015, the evaluator used a market baseline for residential furnaces, air conditioning, and ECM measures, and they are working toward including other measures using a market baseline—possibly for nonresidential lighting depending upon their ability to access sufficient sales data. In some other instances, the evaluator used a billing analysis to derive a net value. These programs include New Homes and Home Performance with ENERGY STAR (HPWES).

Where adequate baseline data is not available or billing analysis was not practical, the SRA methodology was used for both residential and nonresidential NTG estimates. In some cases, a blended approach was used, which combines measure-level results from both the SMP and the SRA to determine weighted average program NTG ratios. To help minimize response bias, the SRA approach is based on a series of several questions instead of a single response. Similar to MA, decision makers are asked what they would have done in absence of the program. "To establish what decision makers might have done in the program's absence, using questions addressing five core FR dimensions for residential programs and six core FR dimensions for nonresidential programs:

- Would participants have installed measures without the program?
- Were participants planning on ordering or installing the measures before learning about the program?
- Would participants have installed the measures at the same efficiency levels without the program incentive?
- Would participants have installed the same quantity of measures without the program?
- In the program's absence, would participants have installed the measures at a different time?

[35] Ibid, page 194.

- Was the purchase of the measures in the organization's most recent capital budget? (Nonresidential only)[36]
    - *Noteworthy Differences from Massachusetts.* Based on responses to each of the FR items, each respondent, who begins with a FR score of 1.0, receives a percentage decrement. The size of the decrement varies, depending on the response, and results in a final FR score between 0 percent and 100 percent. The program FR score is the savings-weighted average of individual respondents' scores. In addition, nonparticipant SO is not addressed in the Focus report or utilized in the NTG estimate using the SRA approach.
    - When using the SMP methodology, net savings is calculated as the difference between the average market baseline and the average program-installed energy consumption, under the assumption that FR is captured in the baseline. The Massachusetts C&I Baseline Framework describes the industry standard practice as "the equipment or practice specific to the application or sector that is commonly installed absent program intervention." Furthermore, the Massachusetts framework states "For commodity measures, all baselines used for evaluated gross savings must be based on market-level research and not depend on a single participant's perspective. Free ridership questions should be based on the premise of this same market-defined baseline. As such there is inherent separation of baseline and free ridership effects that avoids overlap or double counting of any free ridership effect."

2.9.2 Timing

The 2014 Report appendices do not provide any guidance related to timing of measurement.

2.9.3 Baseline

In calendar year 2013, Wisconsin started transitioning away from estimating net savings based exclusively upon the results of SRA surveys to the SMP—an approach that measures the impact of the programs on the average efficiencies of measures sold and installed in Wisconsin. The SMP methodology was applied to nine residential measure categories and four nonresidential measure categories. These measure categories were selected because they had adequate baseline data. In some cases, a blended approach was applied by combining measure-level results from both the SMP and the SRA to determine savings-weighted average program NTG ratios.

[36] Ibid, page 201.

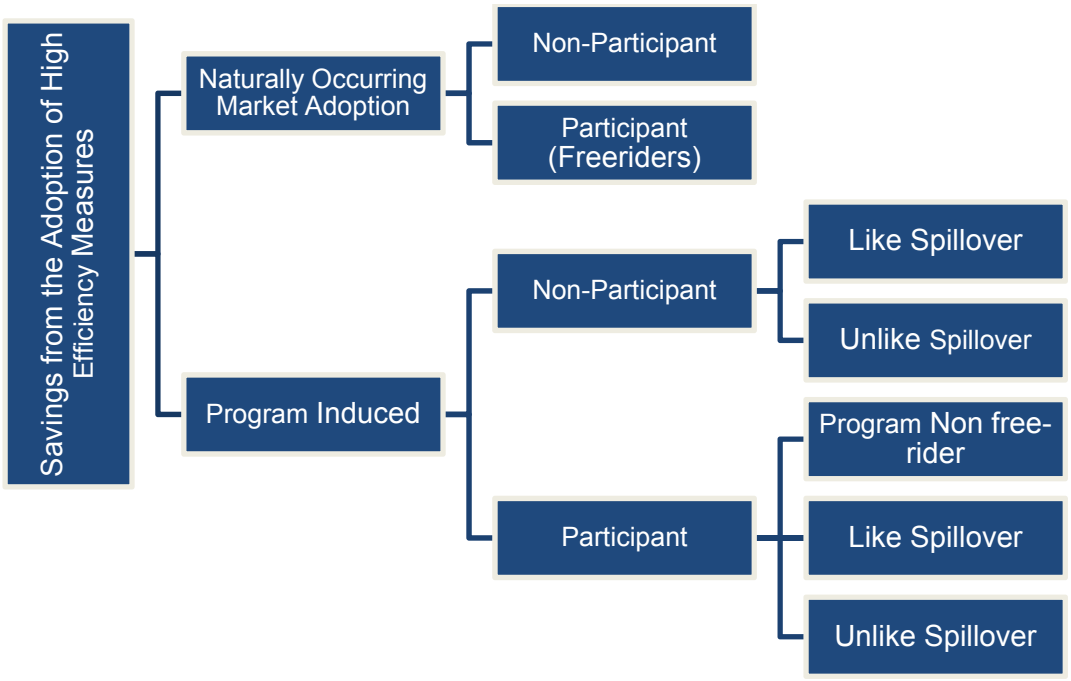
3.0 CONSIDERATIONS FOR INTEGRATING NET-TO-GROSS AND NET SAVINGS METHODS

3.1 Background

The Massachusetts Residential and C&I NTG frameworks[37] allow for many types of NTG methods that capture different components of net savings. Since the publication of the two framework documents in 2011, Massachusetts has expanded on the methods employed in net savings research (see Appendix A for a summary of common net savings estimation methods and components being captured in MA). The goal of this net savings research is to provide the best description of the influence the programs have had on energy savings in Massachusetts, whether by program participants or the larger market.

Depending on the method and scope of evaluation research, program net savings and their components may be captured in isolation or be combined across two or more component categories. The components of net savings are illustrated in Figure 3-1, below.

Figure 3-1. Sources of Non-program and Program Induced Savings



[37] Tetra Tech; KEMA; NMR Group, Inc. (2011). Cross-Cutting (C&I) Free-Ridership and Spillover Methodology Study Final Report. Massachusetts Program Administrators, and NMR Group, Inc. and Tetra Tech (2011). Cross-Cutting Net to Gross Methodology Study for Residential Programs–Suggested Approaches. Final report prepared for the Massachusetts Program Administrators.

Integrating the results of two or more methods can lead to a more complete understanding of program attributable savings. The benefits of a more complete understanding include the following:

- More closely aligning prospective NTG ratios with the full effect of the program across a market or sector.
- Identifying additional energy savings beyond participant-year net savings and incorporating the results into a NTG ratio.
- Informing PAs and program implementers about full program effects in a market and identifying market transformation or opportunities for program designs to capture different sources of net savings. For example, a study finding a program with low market participation but high attribution by participants may be able to drive additional savings by expanding efforts with suppliers, although with some risk of higher self-reported FR by end-users.

The challenges and opportunities of integrating multiple methods relate to the components of net savings that each method captures and the scope of time represented by the methods. These challenges and opportunities include the following:

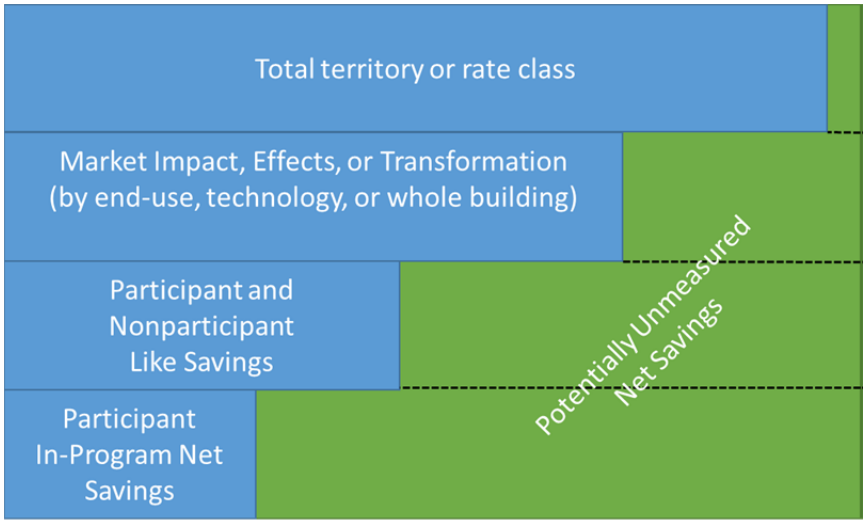
- Avoiding the double counting of impact factors when two or more methods capture the same net savings component
- Improving the accuracy of NTG ratios by triangulating or otherwise combining methods to deliberately identify the same net savings components (e.g., FR)
- Combining the net savings results from studies that may occur with different study participants over time
- Allocating market-level savings to specific programs while avoiding double counting
- For methods requiring a history of data, collecting that data over time and combining it with methods that require a relatively short time period for data collection.

Participant FR and short-term like SO have a well-established history and practice in Massachusetts. The challenge for integration comes when attempting to quantify nonparticipant SO or larger market or sector effects using additional methods, where the scope of analysis of the additional market studies do not necessarily align with those of the FR/SO studies. The current FR/SO work is at the technology group level for C&I programs and program level for residential programs. The scope for data collection and analysis in market studies don't necessarily align with these divisions. Further, market studies tend to include total effects including participant non-FR, as well as possibly overlapping multiple programs. As examples:

- Top-down methods have the potential to identify full sector-level portfolio net savings but cannot allocate savings to specific measures or programs
- Market effects studies may examine measures that cross sectors or programs (e.g., smaller HVAC measures or retrofit vs new construction programs), with sales data that may not be associated with a specific sector or program
- Participant self-report methods may capture FR and short term like-SO, with a market effects study capturing NPSO and participant NTG in its overall net savings results for specific years and over time. The participant self-report NTG results may be a subset of the market effects study's results.

Figure 3-2 illustrates broad categories of net savings studies and the potential for double counting (if multiple methods are used) or failing to count net savings (if a single method is used). For purposes of illustration, Figure 3-2 shows large differences between the scopes of each category and suggests greater net savings may be found with a more broadly-scoped approach (e.g. market effect vs participant self-report). In fact, it is possible that the differences between a broadly-scoped study and a narrowly-scoped study could very small, with potential for the broader study to show lower net savings than the narrow study.

Figure 3-2. Scopes of Potential Net Savings and Research



Evaluation teams, when integrating the results from multiple studies, need to consider several factors when selecting studies to represent various NTG components. These include:

- The scale and scope of each study, including areas of overlap and gaps between the studies
- The age of each study and its relevance to current or prospective conditions
- The quality of each study in terms of how well it captures or specifies NTG components
- The coverage of each study and how it applies to specific programs or measures
- Whether savings should be reported at a program or portfolio level.

These factors should be examined to determine the data to be used to "triangulate" information to arrive at net savings or net-to-gross conclusions. For example, participant self-reported FR may be compared to influential vendor results, and other market information. Although evaluators use expert judgement in combining information sources, when adding the complexity of market-level studies and identifying the components of NTG or net savings that include overlapping findings or gaps, a structured approach may be useful to drive consistency and transparency. This is not currently addressed in the Massachusetts C&I and Residential NTG frameworks.

The Illinois TRM describes one approach that may be adaptable to Massachusetts. In the Illinois TRM, a chapter recommends NTG methodologies and approaches[38] for cases where multiple

[38] Illinois Statewide Technical Reference Manual – Attachment A: IL-NTG Methodologies. IL TRM v5.0 Vol. 4_February 11, 2016_Final, p. 68.

methods are used to determine NTG ratios. The TRM provides guidance that evaluators can use to "triangulate" the results for studies by rating the analysis methods and data on 0 to 10 scale. For each study, the rating is based on three criteria:

1. How likely is this approach to provide an accurate view of free-ridership?
2. How valid is the data collected and analysis performed?
3. How representative is the sample?

The weight for each method is the average score for that method across the three questions, with each studies' contribution to a NTG ratio proportional to its score compared to other studies' scores.

The Illinois approach may be useful for Massachusetts in concept, although the focus on FR and the specific questions being used to weight the studies appear inadequate to address all components of NTG that a study that a study seeking to integrate NPSO or market effects may cover. That said, the concept of weighting or selecting the best study to capture a component of NTG across a series of studies or methods may be applicable to Massachusetts and allow for an explicit and transparent approach to integrating the findings from multiple studies to arrive at a single NTG ratio.

Reporting savings at the program or portfolio level is another issue that may need to be reviewed for Massachusetts. The current practice is to report savings at a program or technology level. The C&I and Residential evaluation frameworks do not address the issue of savings that may not be attributable to specific programs nor do the frameworks describe an approach on when and how those savings should be assigned to a sector's or portfolio's savings. The issue of assigning savings at sector (e.g., C&I or residential) or program portfolio (e.g. all programs) levels may need to be reviewed for Massachusetts as evaluation research considers market-level or nonparticipant spillover effects.

One state, New York, provides an example and precedent to explicitly allow for net savings to be reported at a program portfolio level for "difficult-to-assign" net savings or savings identified at the market-level, rather than program level. In the New York Evaluation Plan Guidance for EEPS Program Administrators,39 the guidance recommends that "net savings resulting from market effects studies must be included in the portfolio-level benefits-costs analysis." The guidance includes in the "difficult-to-assign" category "program effects that cannot be reliably attributed to a specific program…" and would be incorporated into a market effects study that also considers "other non-program effects due to such factors as DOE ENERGY STAR, and the gradual non-program induced evolution of the market in terms of attitudes, knowledge, and behavior regarding energy efficiency." An example in the New York guidance uses stocking practice to illustrate the point – stocking practices may have shifted to favor energy efficient products. These products may not be easily allocated to specific programs and thus net savings due to the stocking practices could be applied at the portfolio level.

[39] *New York Evaluation Plan Guidance for EEPS Program Administrators.* Prepared by the New York State Department of Public Service and the Evaluation Advisory Group – August 2008 Updated August 2013. Appendix F, pp. xv.

3.2 A General Approach to Integrating the Results of Two or More Net Savings Studies

Based on the goals of integrating studies that address net savings, reviewing the recent past evaluation practices of Massachusetts NTG studies, and identifying the considerations for integrating NTG studies, the evaluation team has developed a general approach that can be used by others as they seek to integrate the results of two or more studies. To a degree, this general approach is industry practice in terms of how evaluators think through and combine the results of two or more net savings studies. In presenting this general approach, it is the intent of the evaluation team to create a structure and facilitate discussions around the relative qualities of different studies that may be integrated, as well as facilitating transparency in the factors driving the decisions.

The general approach organizes results from each study in a table that shows findings for each net savings component as well as the qualities or key considerations of each study's results. The details captured are as follows:

- The result of specific NTG components captured in a study or how cross-component (holistic) results combine to inform the specific component. For example, is SO differentiated between participants and nonparticipants, or was only one type of SO researched?
- Overall NTG or net savings results
- Statistical confidence and precision, if available
- Any concerns an evaluator may have about over- or under-stated results, including judgements made with qualitative or subjective information
- Baseline differences between studies.

Evaluators would use the summary information to inform what parts of each study might be used to inform final NTG or net savings results. The tables might be shared within an expert panel to ensure consistent information is being considered by each participant in the panel and may also be used to identify critical gaps in the information that suggest a need for additional specific research.

The following table provides examples of information from two hypothetical studies that an evaluator might include in the tables. The example table shows the information side-by-side to illustrate how information from the two hypothetical studies would be combined to inform NTG results for a prescriptive program. Gray shaded cells represent NTG components not captured for a given study, with blue shaded cells indicating results that are captured in the study, but aggregated with other components.

Table 3-1. Example of Hypothetical Summary Information Used for Integration

Net Savings Components and Considerations	Study 1 – Participant Self Report for prescriptive retrofit rebate program	Study 2 – Market Effects Study using sales data
Naturally Occurring Market Adoption (NOMA)		
Non Participant (NP) Like Measure	Not included	Total Massachusetts efficient market share is 75% based on sales data NOMA is estimated as 25% based on market share of comparison non-program area sales data. Program share of the market is 40%
Participant Free Ridership (P FR)	P FR is estimated as 0.15 based on a program participant survey using 2016 survey and algorithm	
Program Induced Savings		
Participant Non-Free Ridership (P non-FR)	P non-FR = 1 – FR = 1 - 0.15 = 0.85	Not isolated—included in total program effect
Participant Like Spillover (P Like SO)	P Like SO is estimated as 0.05 based on a participant survey conducted 6 to 12 months after program participation	Not isolated—included in total program effect
Participant Unlike Spillover (P Unlike SO)	Not included	Not included
Nonparticipant Like Spillover (NP Like SO)	Not included	Not isolated—included in total program net effect
Nonparticipant Unlike Spillover (NP Unlike SO)	Not included	Not included
Whole-building SO	Not included – not relevant to program	Not included
Results		
Net to Gross (NTG)	NTG = 1 – FR + P SO = 0.85 + 0.05 = 0.90	Total Program Effect as NTG = (Total Massachusetts efficient market share – estimated share absent the program)/(program share) = (75%-25%)/40% = 1.25
Other Considerations		
Statistical precision	90% +/- 10% for NTG result	Not applicable
Concerns about over- or understating results and other judgements	Survey respondents did not include the largest projects – 10 percent of savings. Response rate of 33 percent typical, but non-response bias unaccounted for. Statistical confidence and certainty could result in up to a +/- 10% shift in combined free ridership and like-spillover results. Open ended questions used for consistency check and for adjusting final individual respondent NTG score.	Comparison area sales data estimated to cover half of the market sales, with some risk that unknown sales may show a different efficient market share. Market data in Massachusetts covers 60 percent of the market and does not include full coverage of all brands supported in the program. The missing brands have a small impact on the program and are not .dominant market actors, so not expected to create a large threat to validity. Overall view of market effects and program attributable sales viewed as solid, but differences in sales data and comparison area may understate attributable sales. Sales share of highest efficient equipment much less in comparison area.
Baseline considerations	Participant survey included questions linking NTG to the industry standard practice (ISP) baseline.	Sales data and study covers a time frame prior to shift in industry standard practice (ISP) baseline.

Using the example integration table shown above, an evaluation team ***may*** make the following observations and judgements, simplified for illustration purposes:

1. Study 1 is designed to estimate NTG relative to the ISP but Study 2 is not. Use of the Study 2 results would require adjustment to the ISP baseline.
2. Study 2 is more comprehensive than Study 1, because Study 2 includes NP like SO in program net effects and Study 1 does not.
3. The Study 2 results have greater methodological uncertainty than the Study 1 results because of identified weaknesses in the assumption that the comparison area is a good representation of Massachusetts absent the program. Taking these uncertainties into account, a reasonable lower bound for NTG based on Study 2, without baseline correction, is 1.15.
4. With baseline correction, a reasonable lower bound for NTG based on Study 2 is 1.10, to account for the higher ISP baseline in Massachusetts relative to the comparison state.
5. Self-report NTG results are reasonable for program tracked sales, but the Market data from Study 2 point to larger shifts in the market than is being captured by the Study 1 results.
6. Given the uncertainties identified in each study, the recommendation is [examples of possible recommendations are a, b, or c below].
    a. Use the Study 1 NTG result while recognizing it omits some program effects, because the Study 2 results are too uncertain.
    b. Use the Study 2 NTG lower bound adjusted for ISP baseline.
    c. Conduct additional market actor research in Massachusetts to provide better information on the program effect at shifting the market outside the program.

3.3 Additional Considerations

Evaluators may be faced with a number of combinations of different study types and scopes. Given that studies may have varying quality in regards to capturing NTG components, evaluators must decide how those studies and their quality considerations should be used to inform a combined NTG ratio. These quality considerations include:

- The result of specific NTG components captured in a study or how cross-component (holistic) results combine to inform the specific component. For example, is SO differentiated between participants and nonparticipants, or was only one type of SO researched?
- Overall NTG or net savings results.
- Statistical confidence and precision, if available.
- Any concerns an evaluator may have about over- or under-stated results, including judgements made with qualitative or subjective information.
- Baseline differences between studies.

Below we list several possible scenarios and the general approach used to resolve integration issues.

Scenario 1: Self-report studies combined with market studies

Self-report studies may include participants, nonparticipants, or market actors. In Massachusetts, participant self-reports are used to determine FR and like-measure SO, and are applied to specific programs. In this context, and particularly for C&I programs, influential market actors are surveyed to help determine project-specific net savings. Results are organized by program and used to develop prospective NTG ratios.

Relying solely on self-reports related to specific measures and programs limits the ability of evaluators to recommend a NTG ratio that considers larger market effects or nonparticipant SO. As such, market studies may be useful to combine with current self-report methods to develop a more holistic view of program impacts and/or identify where program designs might be adjusted to capture additional net savings.

Market studies using sales data or trade ally market-level self-reports related to specific measures or end uses may be integrated with participant self-reports with the following guidelines in mind:

- Trade ally market-level self-reports may be useful to inform ways to interpret participant FR, for example, by creating a ceiling on FR or overriding the self-report results. Evaluators should be mindful that the trade ally results at the market level may also be capturing NPSO and potentially market effects that extend beyond just participant FR. The use of trade ally market-level self-reports is different from the use of influential vendor surveys, which focus on specific customer or project transactions, rather than the overall market.

- Market sales data analysis, if it includes high-quality comparison area sales data, can describe total net savings for a program, including nonparticipant SO. However, the results should be weighed against participant and influential vendor self-reports and should be verified by the quantities of measures recorded by the program for rebate processing. Assuming the quality of both the market sales analysis and participant/influential vendor self-reports are high, participant/influential vendor results can be directly compared to understand the net savings related to participant decisions and nonparticipant decisions.

- When faced with multiple data sources of varying quality, such as incomplete market sales data, studies from different time periods, or less reliable self-reports, evaluators should weigh each study relative to its quality at addressing NTG components (e.g., FR, PSO, NPSO). To cover the entire range of possible NTG components, evaluators will likely need to combine market sales, trade ally self-reports, and participant/influential vendor self-reports. The result of the weighting and judgement of information quality can be used to inform the relative certainty regarding net savings components and should be used to assist the PAs and regulatory bodies regarding prospective NTG ratios.

In some cases, programs may cut across equipment categories. For example, new construction or whole-building retrofit programs can affect building shell, heating and cooling systems, and other end use energy efficiency measures such as lighting. In a traditional new construction evaluation, net savings are developed from project-specific analysis via participant or vendor self-reports. Net effects captured by the market sales study should be carefully assigned between equipment programs and new construction programs. Evaluators should keep the

following guidelines in mind when combining cross-market programs with market studies and other equipment-based programs:

- Compare the results and measures being identified in a whole building to the equipment specific market study to reconcile the results assignable to whole-building programs and equipment based programs. End-uses should be carefully reviewed to avoid double counting equipment program net savings in the cross-equipment program. Equipment program sales should first have whole-building program like-measures assigned to the whole building program, with the remainder assigned to the equipment program.
- If the results of multiple studies cannot explicitly account for whole building and equipment program sales, expert judgement will be needed to weigh the results toward each program.
- Net savings market studies that include measures not part of an equipment-based program can be assigned to the whole building programs, although they must also be reconciled between whole-building retrofit and new construction programs, ideally via their individual program evaluations.

Scenario 2: Using market or supply side studies to enhance self-report approaches

In 2008, Wisconsin Focus on Energy's evaluation team issued a memo describing the integration of supply-side study results with end-user NTG self-reports.[40] The memo pointed to uses of supply-side data that could be used to inform or enhance end-user self-reports. Using the market studies to inform or enhance self-report methods differentiates from Scenario 1 in which separate studies with each study's specific NTG or net savings results are integrated with other studies.

Potential integration options include the following:

- *Refine end-user surveys to capture changes identified by a market actor that may influence end-user decision-making.* For example, a market actor might indicate that pricing on energy efficient equipment had changed in response to the program, with end-user self-report research potentially being expanded to test willingness to pay or other price-related indicators of NTG.
- *Inform end-user survey algorithms to capture market changes or practices.* In some cases, end-users may not be aware of all options available in a market or otherwise be influenced by the decision of the market actor. In these cases, the market study can potentially be used to inform specific wording or scoring used in self-report surveys or NTG algorithms to capture the effect of supplier decisions.

[40] Prahl, et al. *Integrating Supply-Side Results With End-User Net-to-Gross Self Reports*. Memo to the Public Service Commission of Wisconsin, March 21, 2008.

Scenario 3: Using top-down results

Top-down methods capture sector- or portfolio-level net savings via the use of econometric models. To date, top-down methods have not been used in Massachusetts to inform NTG ratios or net savings. Their methods and data sources are still in a research and development phase in the industry, with Massachusetts being a leader in this research. Top-down methods may allow for describing a holistic total-market perspective on net savings. Their incremental value for NTG or net savings research has the potential to offer the following benefits:

- Provide high-level confirmation of savings estimated by other methods
- Provide a portfolio-level reconciliation of combined self-report and market study methods.

Top-down study results will not provide granularity to identify equipment specific or end-use specific net savings. Further, the methods capture program attributable behavioral savings and equipment savings without respect to program participation. The Massachusetts Top-Down Working Group is researching and developing top-down methods. The results from the working group may be useful for incorporation into the Massachusetts NTG framework in the future, but at the current time are not ready to inform the framework or be integrated with other NTG or net savings studies.

Scenario 4: Aligning Baseline Conditions across Studies

In 2016, the Massachusetts PAs published a draft report related to standardizing C&I approaches to developing baselines for use in impact evaluations.[41] For NTG savings, the draft report notes the importance of aligning baseline definitions between gross and net savings analysis. In the case of unique measures (generally non-commodity) for which a recognizable market does not exist, the baseline is defined as "the condition that would have existed absent the funded measure." For measures with a recognizable market (generally commodity measures) and not unique applications, the draft report recommends that baselines rely on population-based industry standard practice (ISP) studies to define the baseline conditions.

The draft report notes the need to orient FR questions using the same ISP baseline assumption being used for gross savings. Doing so aligns the verification of gross and net results. For example, if a survey battery references a baseline different from the ISP or does not specify the intended ISP baseline assumption, respondents may provide FR responses that are accurate relative to the question being posed but when used to adjust gross savings are not in alignment with the verified gross results. A lack of alignment may lead to misadjusting gross savings to net savings due to overlapping or double counting FR effects. For example, if gas furnaces are assumed to have an ISP baseline of 85 percent, a NTG study should consider attribution associated with movement from above the 85 percent baseline for free-ridership and would need to consider how to address an individual respondent's response if the respondent indicated that the program contributed to the respondent's decision to upgrade from an 80 percent efficient furnace.

The same condition and misalignment can occur when integrating market studies with end-user self-reports. In any of the above scenarios with the exception of top-down econometric studies, each study may be identifying different points for baselines. As such, it will be important for evaluation researchers to additionally consider the context of baselines across studies when integrating results. As an example, if a market study is investigating the range of sales of a particular product by efficiency and integrating results from participant self-reports, it will be

[41] DNV-GL. Massachusetts Commercial/Industrial Baseline Framework. Draft Report, October 24, 2016.

important to consider the baseline from which those self-reports are reflecting FR, the supply decisions that may drive end-user choices, and the sales data results. The sales data may reflect the entire range of products in a market, but program participating suppliers may limit technology choices to their customers, which may then influence the relative choices end-users make and reflect on when responding to FR questions. Additionally, comparison area sales may or may not reflect a pre-defined ISP used for gross savings or participant FR surveys.

In some cases, it may not be possible to align baseline conditions as part of study design. In these cases, evaluators may need to apply expert judgement or recommend additional research to reduce the opportunity for baselines to be misaligned.

In Appendix B of this document, we present four examples of approaches that are planned or have been conducted for integrating NTG methods to arrive at a total program NTG in Massachusetts for readers seeking more detail.

APPENDIX A: SUMMARY OF COMMON NET SAVINGS ESTIMATION METHODS AND COMPONENTS BEING CAPTURED IN MASSACHUSETTES

In 2003, Tetra Tech staff (as part of PA Consulting Group) developed a methodology for standardized guidelines to quantify net savings for downstream C&I programs[42]. Then, in 2011, the cross-cutting evaluation team developed a pair of detailed studies (frameworks) that provided recommendations for standardized guidelines to quantify net savings for C&I programs (Tetra Tech et al. 2011[43]) and for residential programs (NMR 2011[44]). Since then, cross-cutting studies have recommended methods for the nonresidential new construction market (NMR 2015[45]), C&I lighting market (DNV GL, NMR Group Inc. and Tetra Tech 2015[46]), and HVAC markets (NMR and Tetra Tech, 2014[47]). The recommendations identify possible overlaps among various net savings estimates and recommend approaches to fully understand net savings for the markets.

A.1 NTG AND NET SAVINGS METHODS

Tables in the 2011 frameworks[48] presented common net savings estimation methods and a high-level view of SO components for C&I and residential programs. To serve as a foundation for understanding methods and net savings component overlaps, these tables have been restructured and expanded (Table A-1 and Table A-2 below) to present the methods and provide a more granular view of the components of net savings that are being captured by each method.

Table A-4 and Table A-5 illustrate the common net savings estimation methods, data sources, types of data collected, and the type of analysis used from the residential and C&I Massachusetts framework documents. Additional to the frameworks and shown in the table below are top-down modeling which is currently being explored in Massachusetts in pilot studies, as well as historical tracing which is often used in market effects studies. All of these methods except for Historical Tracing were attempted in Massachusetts between 2012 and 2016.

[42] PA Consulting Group and KEMA (2003), "Standardized Methods for Free-Ridership and Spillover Evaluation—Task 5 Final Report (Revised)", Prepared for National Grid, NSTAR Electric, Northeast Utilities, Unitil, Cape Light Compact, June 16, 2003

[43] Tetra Tech; KEMA; NMR Group, Inc. (2011). *Cross-Cutting (C&I) Free-Ridership and Spillover Methodology Study Final Report.* Massachusetts Program Administrators.

[44] NMR Group, Inc. and Tetra Tech (2011). *Cross-Cutting Net to Gross Methodology Study for Residential Programs – Suggested Approaches*. Final report prepared for the Massachusetts Program Administrators.

[45] NMR Group, Inc. and Tetra Tech (2015). *Recommended Methods for Assessing Market Effects of CI Lighting and Controls Programs.* Final report prepared for the Massachusetts Program Administrators.

[46] DNV GL, NMR Group, Inc. and Tetra Tech (2015). *Recommended Methods for Assessing Market Effects of Non-residential New Construction Programs.* Final report prepared for the Massachusetts Program Administrators.

[47] NMR Group, Inc. and Tetra Tech (2014). *Recommended Methods for Assessing Market Effects of HVAC Programs.* Final report prepared for the Massachusetts Program Administrators.

[48] Table 3-1 in both framework documents.

Table A-1. Common NTG Methods

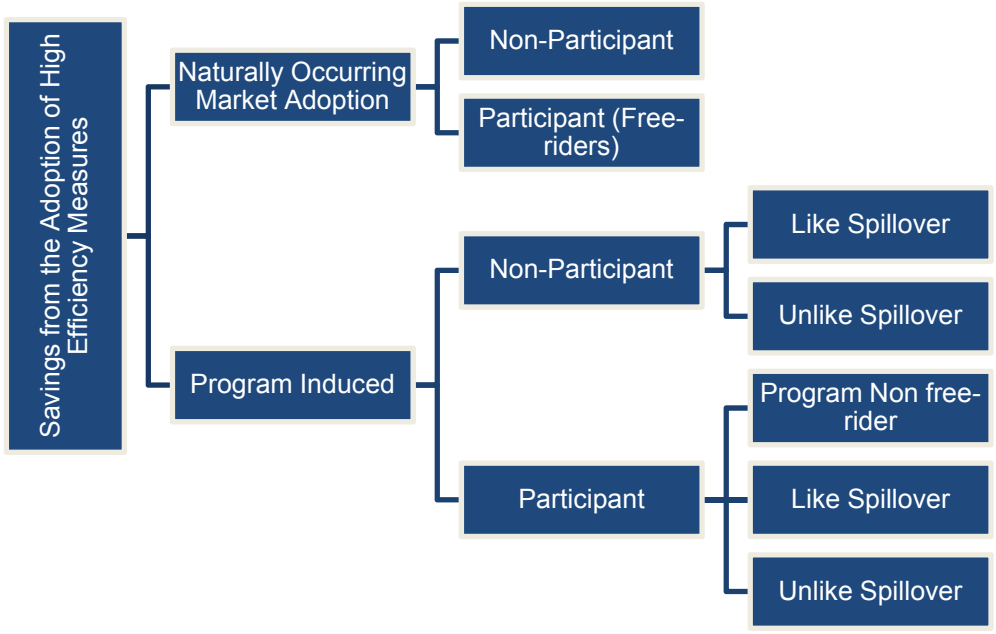
Method	Data Source	Types of Data Collected	Data Description	Analysis	Use in Massachusetts
Market sales data analysis (Cross-sectional studies)	Sales/shipment data provided by industry groups or, ideally, mandated by the federal government	Comprehensive market sales data for program area and comparison area	Sales of efficient and standard equipment in program and non-program areas over time	Weighted/averaged area-to-area comparison, or statistically derived baseline	Residential Lighting CREED Sales Data Analysis
	Manufacturers & regional buyers and distributors	Market sales/shipment data	Sales of efficient and standard equipment in program and non-program areas over time	Weighting and/or averaging	C&I Lighting Market Effects Study (planned)
	Retail store managers and contractors	Sales data	Sales of efficient and standard equipment when program is and is not present	Weighting and/or averaging	
	Retail store managers and contractors, manufacturers, distributors	Self-reported sales or shipments (not sales/shipment data)	Sales of efficient and standard equipment when program is and is not present	Weighting and/or averaging	C&I high bay lighting market effects study – use of contractors
	End-users/decision makers	Self-reported purchases	Self-reported purchases in a specific time period, along with other household behaviors, attitudes, and characteristics, in program and non-program areas	Weighted/averaged area-to-area comparison, or statistically derived baseline	Residential Lighting Multi-State CFL Modeling
	End-users	Panel on-site data collection	Repeat visits to the same buildings to identify new purchases	Weighted/averaged area-to-area comparison	Residential Lighting Panel—MA and NY

Method	Data Source	Types of Data Collected	Data Description	Analysis	Use in Massachusetts
Self-reported counterfactual	Participating and nonparticipating end-users/decision makers	Post Hoc self-reported counterfactual	Self-reported likelihood of buying absent program assistance, and influence of program on purchases outside of program	Scoring and averaging	Many examples—e.g., C&I NTG Surveys
	Retail store managers and contractors, manufacturers, distributors	Post Hoc self-reported counterfactual	Promotional activity and sales with and without program	Weighting and/or averaging	Residential Lighting Supplier Interviews
		Customer-specific influence—self-reported counterfactual	Promotional activity and sales with and without program, among specific customer groups	Weighting and/or averaging	Residential Lighting Supplier Interviews
Pricing and Elasticity Analysis	Nonparticipating end-users	Stated preferences-likelihood	Likelihood of purchase at varying conditions	Scoring and averaging	Residential Lighting
	Participating and nonparticipating end-users	Revealed preferences	Actual purchases, prices, and customer characteristics	Discrete choice analysis or simple average adoption rates	Residential Lighting
	Retail store	Shelf and stocking observations	Observed shelf volumes and prices	Modeling or averaging	Residential Lighting
Billing Data Analysis	Participating customers and comparison group customers (randomized control trials or quasi-experimental design)	Billing data	Consumption data for roughly one year pre- and post-participation, resulting in an estimate of net savings	Weather normalization and change analysis (e.g., a behavior change program using Home Energy Reports)	
Structured Expert Judgment	Experts	Various	NTG estimates from multiple methods, or judging by weight of evidence	Delphi process	Residential Lighting, Residential New Construction

Method	Data Source	Types of Data Collected	Data Description	Analysis	Use in Massachusetts
Top-Down Econometric Modeling	PA, Census and other publicly available data	Aggregate program participation, territory expenditure, and consumption data, demographic / firmographic data. And weather data	Aggregate energy consumption data and data on overall program efforts (e.g., program expenditures) for a large number of cross-sectional observations over a period of multiple years	Time Series/Cross-Sectional regression analysis	Cross-cutting Top-Down Modeling
Historical Tracing (or Case study) Method	Historical data or other information from a variety of sources.	Various	Develop a chronological narrative of the various market influences which can include programs, codes and standards, tax policy and other market interventions	Reconstruction of the events that led to the outcome of interest using weight of evidence; often used in conjunction with Delphi panels to develop estimates of net savings	

The various methods described in Table A-2 estimate different types of program induced and non-program induced energy savings. The different types of savings that may be estimated by different methods are shown in Figure A-1 below and described below the figure.

Figure A-1. Sources of Non-program and Program Induced Savings



1. Naturally occurring market adoption (NOMAD) are savings that are not attributable to programs. These savings come from either nonparticipants who install efficient measures like or unlike program measures on their own, or savings from program FR who would have installed the like efficient measure in the absence of the program.
2. Program-induced participant savings consists of:
    a. Program non FR—savings counted within a program that would not have occurred without the intervention.
    b. Like SO—savings from participants installing additional energy efficient end-use equipment of the same type due to program influences.
    c. Unlike SO—savings from participants installing additional energy efficient end-use equipment of a different type due to program influences.
3. Program-induced nonparticipant savings are those SO savings that occur due to changes in the array of energy-using equipment that manufacturers, dealers, and contractors stock and sell that are causally related to program market interventions.
    a. Like SO—savings from nonparticipants installing energy efficient end-use equipment of the same type due to program influences.
    b. Unlike SO—savings from nonparticipants installing energy efficient end-use equipment of a different type due to program influences.

Table A-2 organizes the methods and the above components of savings that each method captures. Each cell indicates one of the following:

- Isolated (i.e., component separately estimated)
- Accounted for but not isolated (i.e., component is counted in or excluded from the method's estimate of net savings as appropriate, but not separately estimated)
- Blank—method does not address this element or is not well suited to do so.

Table A-2. NTG Components of Savings Estimated by Various Methods

Method	Data Source	Types of Data Collected	Naturally Occurring Market Adoption			Program Induced			
						Participant Savings		Nonparticipant Savings	
			Non-participant	Participant FR	Non-free ridership	Like SO	Unlike SO	Like SO	Unlike SO
Market sales data analysis (Cross-sectional studies)	Sales/shipment data provided by industry groups or, ideally, mandated by the federal government	Comprehensive market sales data for program area and comparison area	Accounted	Accounted	Accounted	Accounted		Accounted	
	Manufacturers & regional buyers and distributors	Market sales/shipment data	Accounted	Accounted	Accounted	Accounted		Accounted	
	Retail store managers and contractors	Sales data records	Accounted	Accounted	Accounted	Accounted		Accounted	
	Retail store managers and contractors, manufacturers, distributors	Self-reported sales or shipments (not sales/shipment data)	Accounted	Accounted	Accounted	Accounted		Accounted	
	End-users/decision makers	Self-reported purchases	Isolated	Isolated	Isolated	Isolated	Isolated	Isolated	Isolated

Method	Data Source	Types of Data Collected	Naturally Occurring Market Adoption		Program Induced				
					Participant Savings			Nonparticipant Savings	
			Non-participant	Participant FR	Non-free ridership	Like SO	Unlike SO	Like SO	Unlike SO
Self-reported counterfactual	Participating and nonparticipating end-users/decision makers	Post Hoc self-reported counterfactual	Isolated	Isolated	Isolated	Isolated	Isolated	Isolated[49]	Isolated
	Retail store managers and contractors, manufacturers, distributors	Post Hoc self-reported counterfactual	Isolated	Isolated	Isolated				
		Customer-specific influence—self-reported counterfactual		Isolated	Isolated				
Pricing and Elasticity Analysis	Nonparticipating or participating end-users	Stated preferences-likelihood	Isolated	Isolated	Isolated	Isolated		Isolated	
	Participating and nonparticipating end-users	Revealed preferences	Isolated	Isolated	Isolated	Isolated		Isolated	
	Retail store	Shelf and stocking observations*							

[49] Using surveys of nonparticipants to capture like or unlike program attributable savings requires careful study design.

Method	Data Source	Types of Data Collected	Naturally Occurring Market Adoption		Program Induced				
					Participant Savings			Nonparticipant Savings	
			Non-participant	Participant FR	Non-free ridership	Like SO	Unlike SO	Like SO	Unlike SO
Billing Data Analysis[50]	Participating customers and comparison group customers	Billing data	Accounted	Accounted	Accounted	Accounted	Accounted		
Structured Expert Judgment	Experts	Various	Isolated	Isolated	Isolated	Isolated		Isolated	
Top-Down Econometric Modeling	PA, Census and other publicly available data	Aggregate program participation, territory expenditure, and consumption data, demographic/ firmographic data. and weather data	Accounted	Accounted	Accounted	Accounted	Accounted	Accounted	Accounted
Historical Tracing	Historical data or other information from a variety of sources.	Various	Isolated	Isolated	Isolated	Isolated		Isolated	

*In and of itself, does not provide net savings estimate.

[50] In general, billing analysis assumes that nonparticipant SO is minimal. If nonparticipant SO is expected to be significant, then a billing analysis may not be an appropriate method. Additionally, depending on the study design billing analysis can also be used for gross savings and may or may not lead to a net savings finding.

Although each of the approaches listed in Table A-2 have the potential for identifying the net savings of various components, specific study designs and results need careful review to minimize or characterize the potential for inaccurate results. Listed below are factors noted by other researchers (NMR et. al. 2011, NEEP 2015[51], UMP[52]) that evaluators should consider regarding a given study's design or results before integrating multiple studies in order to describe how a program may be addressing one or more components of net savings.

Billing Data Analysis

- Billing analysis typically requires up to 12 months of post-implementation consumption data. Generally, 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).
- Nonparticipant SO savings can count against the program to the degree those savings are present in a comparison group. In other words, to the extent that program savings "leak" into the comparison group, they will be assumed to be part of general market trends and not attributed to the program as they should be.
- Billing analysis savings estimates are generally designed to provide aggregate program savings, possibly by subgroup, not individual participant net savings.
- This approach depends heavily on the comparability of the comparison group. Study design should consider whether the comparison group represents net or gross savings. For example, if more recent participants are used as the comparison group, the savings may only represent gross savings.
- The net savings from billing analysis includes, without separating estimate: participant non-FR savings, participant SO, nonparticipant SO, interactive effects, and rebound effects.
- Self-selection can distort the net savings estimates, to the extent the tendency to join the program or not is associated with a tendency to change consumption in other ways. Some analytic methods exist to limit these effects, but outside of randomized treatment control designs the effect cannot be fully eliminated.
- Major decision criteria:
    - Are data available from an appropriate comparison group?
    - Are nonparticipant SO savings expected to be significant?
    - Is the approach designed to minimize self-selection bias?

[51] Johnson Consulting Group (2014). Decision-Framework for Determining Net Savings Approach—Supplemental Document #2 to Principles and Guidance. Prepared for NEEP.

[52] 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.

Market Sales Data Analyses

- Market sales data analyses measures the total net effect of the program, including both FR and participant and nonparticipant "like" SO. 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.
- The results inform a net to gross ratio, which is applied to gross program savings.
- Market sales data analyses are dependent on the availability and quality of sales and shipment data in the area of interest and comparison areas
- Comparison areas may be subject to influence by the program or other programs and should be carefully considered before selecting.
- Impacts are subject to potential bias if market data is limited and without careful selection of a comparison area.
- Major decision criteria: there are a number of important factors to consider when determining if a market sales data analysis is appropriate:
    - Does an appropriate comparison area exist?
    - Are the market data available and complete?
    - What are the features of the program and should a market impact outside of program-tracked project be expected to occur?
    - 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?

Pricing and Elasticity Analysis:

- Many of these approaches rely on customers to identify their intentions, based on actual or stated preferences, compared to hypothetical situations.
- In stated preference models, there is a potential difference between what customers say their purchase intentions are in a hypothetical situation compared to what they actually do. Revealed preference approaches rely on actual purchase activity.
- Pricing and Elasticity models that solely use sales program data and market pricing require adequate variability in prices to identify effects.
- The results inform a net to gross ratio which is applied to gross program savings
- Major decision criteria:
    - How will a survey approach account for the "halo" or "Hawthorne"[53] effects?
    - Is there adequate pricing variability to rely solely on sales and pricing data?

[53] The **Hawthorne Effect** is the process where human subjects of an experiment change their behavior because they are being studied.

- How will this analysis be supplemented with additional approaches to provide context for these finding?
- Are customers likely to be experiencing survey fatigue?

Self-reported counterfactual:

- Despite its drawbacks, this approach is typically the most cost-effective, transparent, and flexible method for estimating NTG. Surveys may target different types of respondents including program participants, program nonparticipants, and market actors.
- End-user self-reported counterfactual is challenging for upstream programs, both because identifying customers who received program benefits is challenging, and because the program tends to be transparent to the customer. When customers can be identified, it is possible to ask them what they'd have done at other price points or if the product had not been available in a particular store. Manufacturers or retailers may provide a better source of information for upstream program.
- The results inform a net to gross ratio which is applied to gross program savings
- Participant surveys only capture a subset of market effects or program attribution.
- Self-reported assessments of program influence are subject to a number of sources of unknown bias. These include:
  - Consumers inability to know what would have done in hypothetical counterfactual situation, the potential for rationalizing past actions or give socially desirable answers, or a general failure to recognize the influence of a program
  - Potential arbitrariness of scoring methods in algorithms.
- Major decision criteria:
  - Is the program or measure type suitable for this approach?
  - What is the potential for response bias and nonresponse bias?
  - How will this analysis be supplemented with additional approaches to provide context for these findings?
  - Are customers likely to be experiencing survey fatigue?

Structured Expert Judgment:

- A particularly useful role for structured expert judging is to develop a consensus estimate to consolidate results from multiple estimation methods or complex data. The Delphi process is the most widely known technique.
- The approach relies on the availability of high-quality data to inform the panel, leading to reasonable estimates of program attribution.
- Sampling-based calculations of confidence and precision are not available.
- Major decision criteria:
  - Is the approach designed to minimize bias among the panel?
  - 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?

Historical Tracing

- 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 approach suggests that if the predicted steps between an activity and an outcome can be confirmed in implementation, this lends a strong argument for causality. Because this method draws from multiple information sources, it is difficult to determine the magnitude of the effects. Rather, it provides evidence that a Delphi panel could consider to quantify a net savings or net to gross ratio estimate.
- 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.
- It may be difficult to translate the influence factors into estimates of impacts without additional modeling.
- Evaluator cannot calculate sample-based statistical confidence and precision levels for the estimate.
- Major decision criteria:
  - Is the historical data available to minimize 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?

Top-Down Econometric Modeling

- 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 program expenditures.
- This approach is subject to the same potential bias as any cross-sectional and/or longitudinal regression analysis due to the influence of confounding variables that are not accounted for in the model. Use of comparison areas can mitigate some of this effect, but can also exacerbate it by introducing additional sources of unaccounted for variation.
- Methods are not fully developed at utility, state or regional levels.
- Relies on high-quality consumption data and data on energy efficiency program efforts over time.
- Top-down methods provide savings at the portfolio or sector level and do not provide savings at the measure, technology or program level.
- The methods provide average savings factors over several years included in the study period, such as savings per dollar of spending or per bottom-up estimated MWh saved. They do not provide separate factors of this type by program year.

- The net savings from the top-down analysis include, but do not separately estimate: participant non-FR savings, PSO, NPSO, interactive effects, and rebound effects.
- Depending on the level of geographic aggregation used, cross-unit SO can count incorrectly against the net savings.
- Depending on the level of geographic aggregation used, participant self-selection can distort the net savings estimates if not accounted for analytically.
- Does not provide information on how to improve program design and implementation and shows savings only after a sufficient number of years have passed with continuous program portfolio operations.
- Major decision criteria:
    - Is there adequate available data to understand program expenditures and energy consumption?
    - Can lag effects be modeled with the available time span of available data?
    - How will this analysis be supplemented with additional approaches, e.g., program-specific process and/or impact evaluations, to provide context for these findings?

A.2 SUGGESTED NTG METHODS BEING USED IN MASSACHUSETTS

This section draws on Table 3-4 in the 2011 C&I[54] and Residential[55] Massachusetts framework documents (both documents contain relevant information in their respective tables numbered Table 3-4).

A.2.1 Residential Programs

The Tetra Tech team reviewed 18 Massachusetts residential studies with NTG components that were posted between 2012 and 2016. These studies encompassed seven programs/program types: Home Energy Service, Multifamily Retrofit, Residential Heating and Cooling, Residential Behavioral, Residential Lighting, and Residential and Nonresidential New Construction. The studies included 15 program evaluations, one study on Top-down modeling, two studies that recommended methods for assessing market effects for the HVAC and Nonresidential New Construction Programs, and one study on overall approaches to measuring market effects of energy efficiency programs. Table A-3 shows the NTG methods used in the 18 program evaluations that were posted between 2012 and 2016.

Table A-3. NTG Approaches for Massachusetts Programs used in 2012-2016 Residential Evaluations*

NTG Approach	Number of Studies
Self-reported counterfactual— end-users/decision makers	7

[54] Tetra Tech; KEMA; NMR Group, Inc. (2011). *Cross-Cutting (C&I) Free-Ridership and Spillover Methodology Study Final Report.* Massachusetts Program Administrators.

[55] NMR Group, Inc. and Tetra Tech (2011). *Cross-Cutting Net to Gross Methodology Study for Residential Programs – Suggested Approaches*. Final report prepared for the Massachusetts Program Administrators.

NTG Approach	Number of Studies
Billing data analysis	4
Others	4
Self-reported counterfactual--supply-side market actors	3
Market sales data analysis—sales/ shipment data (if available)	3
Market sales data analysis—self-reported	2
Structured expert judging	2
Pricing and elasticity analysis--revealed preferences	0
Pricing and elasticity--shelf and stocking survey (w elasticity analysis)	0
Total	18*

*Note some studies include more than one method

- The evaluations most frequently drew upon self-reported counterfactuals from end-users/decision makers. These approaches were used from 2012 to 2015 for programs that included Home Energy Services, Multifamily Retrofit, Cool Smart/HEHE, Residential New Construction, and LED Market Effects Baseline.
- Of the four studies using “Other” methods, two studies were for the Residential Lighting Program from 2015 (demand elasticity using program data, changes in socket saturation, web-based pricing model); one study was the Top-down modeling from 2015; and one study was for the Residential New Construction program from 2014 (compare REM/Rate models for as-built to models with counterfactual assumptions).
- The self-reported counterfactual--supply-side market actors were used for the Residential Heating, Water Heating, and Cooling Equipment Evaluation from 2013, and the Multistage Lighting Net-to-Gross Assessment and LED Market Effects Baseline from 2015.
- The self-reported market sales data analyses were used for the CFL Market Lift study from 2014, and the Multistage Lighting Net-to-Gross Assessment from 2015.
- The 2015 Residential Lighting Program was noteworthy in its utilization of a wide range of NTG measurement approaches including market sales data analysis (shipment and self-reported data), self-reported counterfactuals from market actors, demand elasticity analysis, changes in socket saturation, web-based pricing model, and structured expert judging. More discussion about the integration of these multiple methods is discussed in detail later in this document.
- The Residential New Construction Net Impacts study from 2014 also was noteworthy in its utilization of NTG measurement methods that were not suggested for this program in the 2011 framework document. This study drew upon self-reported counterfactual from market actors and structured expert judging to compare REM/Rate models for as-built to models with counterfactual.
- The Residential Heating, Water Heating, and Cooling Equipment Evaluation from 2013 also notably used multiple NTG measurement approaches including market sales data

analysis (shipment and self-reported data) and self-reported counterfactual from market actors.

- Of the four billing analyses, three were for the Opower Program from 2013 to 2016 and one was for the Home Energy Services Impact Evaluation from 2012.
- The 2011 residential framework recommended using demand elasticity analyses from either revealed preferences or shelf stocking data for the ENERGY STAR Lighting Program. However, the 2015 Lighting evaluation instead drew upon program data to perform demand elasticity analysis. Since this relied on a different data source from the revealed preferences and shelf stocking survey that were recommended in the 2011 residential framework, it is shown in the "Others" column.
- Pricing and elasticity analyses (revealed preferences and shelf stocking survey) were also recommended in the 2011 residential framework for the ENERGY STAR Appliances program. As there were no published evaluations in the 2012-2016 period for the Appliances program, Table A-4 shows that none of the evaluations utilized these approaches.

Table A-4 summarizes the residential studies examined. The color coding in each cell represents whether a framework recommended method was used in a given study, whether a method was used but not suggested in the framework, or whether the method was applied but not used for NTG estimation as follows:

Framework recommended method used	Method used but not suggested in the 2011 framework for this program	Method used but not applied to NTG or ME analysis

References for the residential study can be found in Appendix C.

Table A-4. NTG Approaches for Massachusetts Residential Programs used in 2012-2016 Evaluations*

Key to Methods Used in 2012-2016 Evaluations*	Framework recommended method used	Method used but not suggested in the 2011 framework for this program	Method applied but not used for NTG estimation

Program	Technologies Included	Market sales data analysis—sales/ shipment data (if available)	Market sales data analysis—self-reported	Self-reported counterfactual--supply-side market actors	Self-reported counterfactual— end-users/decision makers	Pricing and elasticity analysis--revealed preferences	Pricing and elasticity--shelf and stocking survey (elasticity analysis)	Structured expert judging	Billing data analysis	Others
Home Energy Services Net-to-Gross Evaluation (Completed in 2012)	CFL, air sealing, insulation, refrigerator	Program data used to supplement discrete choice analysis			Direct self-report (participant FR, participant and nonparticipant SO) and discrete choice analysis					
2012 Home Energy Services Impact Evaluation. (Completed in 2012)	CFL, air sealing, furnaces, boilers, DHW, duct treatment, insulation, refrigerator								Billing analysis--fixed-effects, conditional savings analysis (CSA), paired-months modeling	

Program	Technologies Included	Market sales data analysis—sales/ shipment data (if available)	Market sales data analysis—self-reported	Self-reported counterfactual--supply-side market actors	Self-reported counterfactual— end-users/decision makers	Pricing and elasticity analysis--revealed preferences	Pricing and elasticity--shelf and stocking survey (elasticity analysis)	Structured expert judging	Billing data analysis	Others
Massachusetts Multifamily Program Impact Analysis. (Completed in 2012)	Comprehensive retrofits. Residential measures for units and C&I measures for common areas				FR and PSO from property managers and tenant surveys. Informed program planning (no NTG or net savings estimate)					
Wi-Fi Programmable Controllable Thermostat Pilot Program Evaluation. (Completed in 2012)					NTG based on FR from participant survey. Only for process evaluation (no net savings estimate)					
2012 Home Energy Services Pre-Weatherization Initiative Evaluation. (Completed in 2013)	Pre-weatherization initiative				Participant FR					

Program	Technologies Included	Market sales data analysis—sales/ shipment data (if available)	Market sales data analysis—self-reported	Self-reported counterfactual--supply-side market actors	Self-reported counterfactual— end-users/decision makers	Pricing and elasticity analysis--revealed preferences	Pricing and elasticity--shelf and stocking survey (elasticity analysis)	Structured expert judging	Billing data analysis	Others
2012 Residential Heating, Water Heating, and Cooling Equipment Evaluation: Net-to-Gross, Market Effects, and Equipment Replacement Timing. (Completed in 2013)	HVAC. Gas furnaces, boilers, and water heaters		Distributors sales data to assess Net Market Effects	FR and SO from contractor surveys	FR from participant surveys					
Massachusetts Cross-Cutting Behavioral Program Evaluation Integrated Report. (Completed in 2013); *Massachusetts Cross-Cutting Behavioral Program Evaluation Opower Results Memo (Completed in 2015);* *Berkshire Gas Home Energy Report Program Evaluation (Completed in 2016)*	Behavioral Program								Billing analyses	

Program	Technologies Included	Market sales data analysis—sales/ shipment data (if available)	Market sales data analysis—self-reported	Self-reported counterfactual--supply-side market actors	Self-reported counterfactual— end-users/decision makers	Pricing and elasticity analysis--revealed preferences	Pricing and elasticity--shelf and stocking survey (elasticity analysis)	Structured expert judging	Billing data analysis	Others
Massachusetts Residential New Construction Net Impacts Report. (Completed in 2014)	Building practices for shell, HVAC, water heating, lighting				Factors influencing changes to homebuilder practices between 2004 and 2011			EE values in absence of program for 12 measures in participating and nonparticipating homes		Compare REM/Rate models for as-built to models with counterfactual assumptions. NTG ratios (FR and NPSO) calculated by comparing combined net savings of program and nonprogram homes to estimated savings claimed for 2011 program year
Market Lift Assessment. (Completed in 2014)	CFLs	NTG based on sales above pre-lift period								

Program	Technologies Included	Market sales data analysis—sales/ shipment data (if available)	Market sales data analysis—self-reported	Self-reported counterfactual--supply-side market actors	Self-reported counterfactual— end-users/decision makers	Pricing and elasticity analysis--revealed preferences	Pricing and elasticity--shelf and stocking survey (elasticity analysis)	Structured expert judging	Billing data analysis	Others
Massachusetts Upstream Lighting Program Net-to-Gross Ratio Estimates Using Supplier Self-Report Methodology (Completed in 2015)	CFLs, LEDs			IDIs with lighting manufacturers, retail lighting buyers; and surveys of store manager						
Preliminary Lighting Demand Elasticity Findings Memo (Completed in 2015)	CFLs									Demand elasticity analysis using program data
MA Residential Point-of-Sale Modeling (Completed in 2015)	CFLs, LEDs	CREED and Program data								
Multistage Lighting Net-to-Gross Assessment (Completed in 2015)	CFLs, LEDs, Fixtures		Self-reported end-user purchases in Massachusetts and comparison states					Consensus approach (PAs, EEAC consultants, evaluation team)		Changes in socket saturation; web-based pricing model (context for prospective NTG estimates)

Program	Technologies Included	Market sales data analysis—sales/ shipment data (if available)	Market sales data analysis—self-reported	Self-reported counterfactual--supply-side market actors	Self-reported counterfactual— end-users/decision makers	Pricing and elasticity analysis--revealed preferences	Pricing and elasticity--shelf and stocking survey (elasticity analysis)	Structured expert judging	Billing data analysis	Others
Final Draft Report of Massachusetts LED Market Effects: Baseline (Completed in 2015)	LEDs		End-user self-reported purchases and supply side actor self-reported market share				Shelf stocking survey to understand availability and pricing (no elasticity analysis			
Top-down Modeling Methods Study (Completed in 2015)	All electric end uses									Top-Down Modeling
Recommended Methods for Assessing Market Effects of HVAC Programs (Completed in 2014)[56]	Central ACs and heat pumps, Gas furnaces and boilers, Ductless mini-split heat pumps	HARDI data	Distributor panel	Participating and nonparticipating contractor survey	Participating and nonparticipating customer survey					
Recommended Methods for Assessing Market Effects of Nonresidential New Construction Programs (Completed in 2015)[50]	All building practices		Non-residential new construction baseline study		Gas & Electric NTG Studies			Attribution between programs		Compare as-built models to counterfactual based models

[56] These studies recommended approaches for identifying market effects, but do not represent actual evaluations.

Program	Technologies Included	Market sales data analysis—sales/ shipment data (if available)	Market sales data analysis—self-reported	Self-reported counterfactual--supply-side market actors	Self-reported counterfactual— end-users/decision makers	Pricing and elasticity analysis--revealed preferences	Pricing and elasticity--shelf and stocking survey (elasticity analysis)	Structured expert judging	Billing data analysis	Others
Methods for Measuring Market Effects of Massachusetts Energy Efficiency Programs (Completed in 2015)[50]	General	Cross-sectional analysis	Cross-sectional analysis	Interviews with retail store managers, contractors, distributors, manufacturers				Various		Forecasting or retrocasting non-intervention baseline

A.2.2 C&I Programs

Table A-5 summarizes the 12 C&I studies examined. The color coding in each cell represents whether a framework recommended method was used in a given study, whether a method was used but not suggested in the framework, or whether the method was applied but not used for NTG estimation or market effects analysis.

Method used	Method used but not suggested in the 2011 framework for this program	Method used but not applied to NTG or ME analysis

References for the C&I studies can be found at the end of this appendix.

Table A-5. NTG Approaches Used for in Specific C&I Evaluation Studies in Massachusetts

Key to Methods Used in 2012-2016 Evaluations*	Framework recommended method used	Method used but not suggested in the 2011 framework for this program	Method applied but not used for NTG estimation

Study	Technologies Included	Market sales data analysis—sales/ shipment data (if available)	Market sales data analysis—self-reported	Self-reported counterfactual --supply-side market actors	Self-reported counterfactual—end-users/decision makers	Others
High-Bay Lighting Market Effects (Completed in 2011)	*High-Bay Lighting*		Market actor and end user self-reports	Participating/Non-participating: Awareness of programs and influence of programs	Participating/Non participating: Awareness of programs and influence of programs	Tracking data analysis
LED Market Effects Baseline (Completed in 2015)	*Commercial and Residential LED Products*		End-user self-reported purchases and supply side actor self-reported market share			Shelf stocking survey to understand availability and pricing
C&I LED Spillover (Completed in 2015)	*LEDs, primarily screw-based*				Participating/non-participating	Verified installations from on-site visits and participation data from tracking data
C&I Lighting Market Effects	*HPT8s, Screw-based LEDs, Linear LEDs, Advanced Lighting Controls*	Massachusetts and Comparison Area Distributor total sales data	Contractor self-reports	Participating/Non participating: Awareness of programs and influence of programs		Verified installations from on-site visits and participation data from tracking data
Upstream Lighting (Bright Opportunities Program) NTG	*LEDs and HPT8s*		Market actor and end user self-reports	Participants	Participating contractors and distributors and nonparticipant distributors	
C&I Code Compliance	New Construction			Awareness of codes and PA programs	Awareness of codes and PA programs	On-site verification

Study	Technologies Included	Market sales data analysis—sales/ shipment data (if available)	Market sales data analysis—self-reported	Self-reported counterfactual --supply-side market actors	Self-reported counterfactual—end-users/decision makers	Others
Top-Down Modeling (Completed in 2015)	All electric end uses					Top-Down Modeling
C&I Natural Gas Programs Free-ridership and Spillover Study	All natural gas end uses			Participants	Design professionals and equipment vendors	
C&I Electric Programs Free-ridership and Spillover Study	All electric end uses			Participants	Design professionals and equipment vendors	
Upstream HVAC Market Characterization (In Progress)	Air Cooled Unitary and Split AC Systems, Large commercial RTU and Split Systems, Water and Evaporatively Cooled CAC, ASHP, WSHP, GSHP, Ductless Mini-Split HP,	Total sales by equipment type from distributors				
Boiler Market Characterization Study (In Progress)	Boilers	Review of program and public data to determine market share		Influence of program on sales		
Incentivized High Efficiency Market Share (In Progress)	Lighting, HVAC, Domestic Hot water, EMS					Share of high efficiency equipment that is incentivized based on-site and tracking data analysis

Table A-6 summarizes the C&I studies referenced in Table A-5 around technology types rather than programs. This table shows the recent history of different NTG methods being applied to different technologies. The shaded cells also identify the year of the study if completed or note whether the study is in progress or planned.

Table A-6. NTG Approaches Used for Specific C&I Technologies in Massachusetts

Key to Methods Used in 2012-2016 Evaluations*	Framework recommended method used	Method used but not suggested in framework for this program	Method applied but not used for NTG estimation

Study	Technologies Included	Market sales data analysis—sales/ shipment data (if available)	Market sales data analysis—self-reported	Self-reported counterfactual--supply-side market actors	Self-reported counterfactual — end-users/decision makers	Others
Lighting	All Lighting		Process Evaluation of the 2012 Bright Opportunities Program (2013)			
	High Bay Lighting		HBL Market Effects Study (2011)	HBL Market Effects Study (2011)		HBL Market Effects Study (2011)
	LEDs Only		LED Market Effects: Baseline Characterization (2015)		Massachusetts LED Spillover Analysis (2015)	LED Market Effects: Baseline Characterization (2015) and Massachusetts LED Spillover Analysis (2015)
	Lighting + Controls	Commercial and Industrial Lighting and Controls Market Effects Study (In Progress)	Commercial and Industrial Lighting and Controls Market Effects Study (In Progress)	Commercial and Industrial Lighting and Controls Market Effects Study (In Progress) and 2013 Commercial and Industrial Electric Programs Free-ridership and Spillover Study (2015)	2013 Commercial and Industrial Electric Programs Free-ridership and Spillover Study (2015)	Commercial and Industrial Lighting and Controls Market Effects Study (In Progress) and Top-down Modeling Methods Study (2015)

Study	Technologies Included	Market sales data analysis—sales/ shipment data (if available)	Market sales data analysis—self-reported	Self-reported counterfactual--supply-side market actors	Self-reported counterfactual — end-users/decision makers	Others
HVAC	All HVAC	Upstream HVAC Market Penetration Data Collection (In Progress)		2014-2015 Commercial and Industrial Natural Gas Programs Free-ridership and Spillover Study (2015) and 2013 Commercial and Industrial Electric Programs Free-ridership and Spillover Study (2015)		Top-down Modeling Methods Study (2015) and Assessment of the Share of Incentivized High Efficiency Equipment (Planned)
	Boilers Only	Upstream HVAC Market Penetration Data Collection (In Progress)		Boiler Market Assessment Phase II (In Progress)		
Water Heating	All			2014-2015 Commercial and Industrial Natural Gas Programs Free-ridership and Spillover Study (2015) and 2013 Commercial and Industrial Electric Programs Free-ridership and Spillover Study (2015)		Top-down Modeling Methods Study (2015) and Assessment of the Share of Incentivized High Efficiency Equipment (Planned)
EMS	All			2014-2015 Commercial and Industrial Natural Gas Programs Free-ridership and Spillover Study (2015) and 2013 Commercial and Industrial Electric Programs Free-ridership and Spillover Study (2015)		Top-down Modeling Methods Study (2015 and Assessment of the Share of Incentivized High Efficiency Equipment (Planned)
Refrigeration	All			2014-2015 Commercial and Industrial Natural Gas Programs Free-ridership and Spillover Study (2015) and 2013 Commercial and Industrial Electric Programs Free-ridership and Spillover Study (2015)		Top-down Modeling Methods Study (2015) and Assessment of the Share of Incentivized High Efficiency Equipment (Planned)
C&I New Construction	All			Code Compliance Baseline Study (2012)		Code Compliance Baseline Study (2012)

Study	Technologies Included	Market sales data analysis—sales/ shipment data (if available)	Market sales data analysis—self-reported	Self-reported counterfactual--supply-side market actors	Self-reported counterfactual — end-users/decision makers	Others
Other	Food service, comprehensive, building envelope, thermostats, custom, compressed air, process, steam traps, motors and drives			2014-2015 Commercial and Industrial Natural Gas Programs Free-ridership and Spillover Study (2015) and 2013 Commercial and Industrial Electric Programs Free-ridership and Spillover Study (2015)		Top-down Modeling Methods Study (2015)

APPENDIX B: EXAMPLES OF INTEGRATING STUDIES AND METHODS IN MASSACHUSETTS

This appendix presents four examples of brief case studies for how evaluation teams have integrated or plan to integrate information sources to develop NTG estimates in Massachusetts. The first two examples are forward looking and reflect recommendations on approaches to the new construction markets. The second two examples relate to C&I lighting markets and were used to develop formal NTG recommendations.

B.1 EXAMPLE 1: NONRESIDENTIAL NEW CONSTRUCTION NTG

In February 2015, the cross-cutting evaluation team submitted a document detailing the recommended methods for assessing market effects on nonresidential new construction programs.[57] The objective of that document was to outline appropriate methods for evaluating the programs' effects on this market based on the range of methods available for measuring market effects. This includes methods for establishing qualitative evidence of the programs' effects on markets, and quantifying the effects (which incorporate spillover) as well as estimating net savings.

As discussed in the recommended methods document, multiple studies have been conducted or are proposed to be conducted to estimate one or more components of net savings in the nonresidential new construction market. The recent or proposed studies that provide new construction-specific net savings estimates include the following:

- Commercial and Industrial Electric Programs Free-ridership and Spillover Study[58]
- C&I Natural Gas Programs Free-ridership and Spillover Study[59]
- Final Report of Massachusetts LED Spillover Analysis[60]
- C&I Lighting and Controls Market Effects Study (P53, in progress)
- HVAC market effects evaluation[61]
- Nonresidential New Construction Market Effects Study.[62]

In addition to these studies, the cross-cutting evaluation team is also planning an attribution study as part of their evaluation of the Code Compliance Support Initiative (CCSI) for both the residential and commercial sectors in 2017.

---

[57] Recommended Methods for Assessing Non-Residential New Construction Market Effects (http://ma-eeac.org/wordpress/wp-content/uploads/Recommended-Methods-for-Assessing-Non-residential-New-Construction-Market-Effects.pdf). This report discusses in more detail the approach to accounting for overlaps in net savings estimates.

[58] http://ma-eeac.org/wordpress/wp-content/uploads/CI-Electric-Programs-Free-Ridership-and-Spillover-Study.pdf.

[59] http://ma-eeac.org/wordpress/wp-content/uploads/CI-Natural-Gas-Programs-Free-ridership-and-Spillover-Study.pdf.

[60] http://ma-eeac.org/wordpress/wp-content/uploads/MA_LED_Spillover_FINAL-Report_092415.pdf.

[61] http://ma-eeac.org/wordpress/wp-content/uploads/Recommended-Methods-for-Assessing-Market-Effects-of-HVAC-Programs.pdf – proposed.

[62] http://ma-eeac.org/wordpress/wp-content/uploads/Recommended-Methods-for-Assessing-Non-residential-New-Construction-Market-Effects.pdf – proposed.

Table B-1 below shows the components of net savings that are estimated in each study (marked with a ●). This table highlights the potential overlaps in NTG estimates in the nonresidential new construction market.

Table B-1. Net Savings Categories and Coverage by Existing and Planned Studies—Nonresidential New Construction

Programs Covered	C&I New Construction	Retrofit	Upstream, C&I New Construction, and Retrofit	C&I New Construction, CCSI, Retrofit, and Upstream
Net Savings Studies	Measured by FR and SO Studies (Gas & Electric—2015)	Measured by FR and SO Studies (Gas & Electric—2015)	Measured by Technology-Specific Spillover and Market Effects Studies (LEDs—2015; Lighting and Controls—current; HVAC—proposed)*	Measured by Non-residential New Construction Market Effects Study (proposed)
Savings Not Attributable to Program(s)—Baseline or Counterfactual				
Naturally Occurring Outside Program—Unlike				●
Naturally Occurring Outside Program—Like			●	
Free-Ridership (Naturally Occurring Inside Program)	●			
Savings Attributable to Program(s)				
Inside Program Non-Freeridership	● (b)		● (c)	● (a)
Inside Like Spillover	● (d)			
Outside Like Spillover	● (e)	● (f)		
Nonparticipant Like Spillover				
Outside Unlike Spillover				
Inside Unlike Spillover				
Nonparticipant Unlike Spillover				
Whole-building Spillover				

*There was a High-Bay Lighting Market Effects study completed in 2011, but the results are likely too old to be applicable to future NTG estimates.

Since there are clear overlaps, the PAs will need to calculate and subtract these estimates either from the overall net savings estimates produced by the above studies, or from the net savings estimate produced by the proposed Nonresidential New Construction Market Effects Study.[63] The formula is as follows:

x = a - b - c - d - e - f

where:

x = additional energy savings from nonresidential new construction market effects not already estimated by another study

a = overall market-level savings in nonresidential new construction attributable to the PA programs

b = net-of-free ridership in-program savings stemming from the C&I New Construction Program

c = new construction technology-specific spillover

d = inside like spillover stemming from the C&I New Construction Program (technologies not covered by c)

e = outside like spillover stemming from the C&I New Construction Program (technologies not covered by c)

f = outside like spillover in new construction stemming from the Retrofit Program (technologies not covered by c)

B.2 EXAMPLE 2: RESIDENTIAL NEW CONSTRUCTION NTG

Similar to nonresidential new construction, the residential new construction sector has two studies that estimate overall residential new construction net savings:

- Residential New Construction Net Impacts Report[64]
- Residential New Construction Market Effects Study[65] (proposed).

The primary difference between the two studies is that the latter is designed to estimate net savings from the Code Compliance Support Initiative (CCSI) as well as the Residential New Construction Program (RNC), while the former – conducted before CCSI was implemented – estimated net savings from the RNC Program only. The net savings estimates from these two studies include savings stemming from the following programs:

- Cool Smart and High Efficiency Heating and Water Heating (HEHE), as reported in the *Residential Heating, Water Heating, and Cooling Equipment Evaluation*[66], and as planned in the proposed HVAC Market Effects Study.[67]

[63] Recommended Methods for Assessing Non-Residential New Construction Market Effects (http://ma-eac.org/wordpress/wp-content/uploads/Recommended-Methods-for-Assessing-Non-residential-New-Construction-Market-Effects.pdf). This report discusses in more detail the approach to accounting for overlaps in net savings estimates.

[64] http://ma-eeac.org/wordpress/wp-content/uploads/Residential-New-Construction-Net-Impacts-Report-1-27-14.pdf.

[65] see the Amended Research Plan for 2016 – Cross-Cutting Code Compliance Support Initiative, February 24, 2016).

- Upstream Lighting, as reported in the *Multistage Lighting Net-to-Gross Assessment.*[68]
- Home Energy Services Program (HES) – On-site assessments conducted for the *Residential New Construction Baseline Study* reveal that some newly constructed homes have had their light bulbs replaced through the HES Program. There has been no recent NTG study conducted for the HES Program.

Table B-2 shows overlaps in previous or proposed NTG estimates in the residential new construction market. The best way to specify these overlaps for HVAC and HES might be to identify new homes in the program tracking data. To identify overlaps between the RNC and Upstream Lighting, the PAs might rely on self-reports of home owners whose homes are visited for the residential new construction baseline study, combined with a survey of builders. Please note there will be overlap between net savings estimates produced by the *Residential Heating, Water Heating, and Cooling Equipment Evaluation* and the proposed HVAC Market Effects Study.

[66] http://ma-eeac.org/wordpress/wp-content/uploads/2012-Residential-Heating-Water-Heating-and-Cooling-Equipment-Evaluation_Net-to-Gross-Market-Effects-and-Equipment-Replacement-Timing-Volume-I-June-2013.pdf.
[67] http://ma-eeac.org/wordpress/wp-content/uploads/Recommended-Methods-for-Assessing-Market-Effects-of-HVAC-Programs.pdf.
[68] http://ma-eeac.org/wordpress/wp-content/uploads/Multistage-Lighting-Net-to-Gross-Assessment-Overall-Report.pdf.

Table B-2. Net Savings Categories and Coverage by Existing and Planned Studies—Residential New Construction

Programs Covered	Cool Smart and HEHE	Cool Smart and HEHE	Upstream Lighting	Residential New Construction, CCSI, Lighting, HES, and Cool Smart and HEHE
Net Savings Studies	**Measured by Heating, Water Heating, and Cooling Equipment Evaluation (2013)**	**Measured by HVAC Market Effects Study (proposed)**	**Measured by Multistage Lighting NTG Study (2015)**	**Measured by Residential New Construction Net Savings Study (2014) and Market Effects Study (proposed)**
Savings Not Attributable to Program(s)—Baseline or Counterfactual				
Naturally Occurring Outside Program—Unlike				•
Naturally Occurring Outside Program—Like		•	•	
Free-Ridership (Naturally Occurring Inside Program)	•			
Savings Attributable to Program(s)				
Inside Program Non-Free-Ridership	•	•	•	•
Inside Like Spillover	•			
Outside Like Spillover	•			
Nonparticipant Like Spillover				
Outside Unlike Spillover				
Inside Unlike Spillover				
Nonparticipant Unlike Spillover				
Whole-building Spillover				

New-construction specific net savings from the above programs should be identified, and subtracted either from the overall net savings estimates produced by the associated studies (except for HES, since there has been no such study) or from the RNC Net Savings Study or the proposed RNC Market Effects Study. Going forward, similar studies should seek to identify savings associated with the residential new construction market. Doing so will allow for developing specific metrics for new construction program NTG and avoid including new construction NTG in other programs' results.

B.3 EXAMPLE 3: MASSACHUSETTS LED PROGRAM SPILLOVER STUDY

In the Massachusetts LED spillover study, the evaluation team evaluated SO and FR associated with all non-residential LED lamps supported by PA programs from 2011-2014. The study combined results from the 2013 Commercial and Industrial Electric Programs Free-ridership and Spillover Study with the Massachusetts C&I Market Characterization On-Site Assessments and Market Share and Sales Trends Study and a review of program tracking data. Analysts used the following methods to collect data for the NTG analysis:

- **On-site data collection.** Data from an on-site survey of a sample of commercial customers to estimate the extent of LED lighting purchases during the study period. With these data they were able to obtain verified information on the quantity and technical characteristics of the LED lighting installed. This data, combined with deemed savings from the Massachusetts TRM for each LED product evaluated, provided the team with gross savings at the customer level. The weights developed for the on-site study were also used to for the SO analysis.
- **Tracking Data Review.** Program tracking data review to identify participation status of LED owners in the on-site sample and to determine, with a fairly high level of certainty, which installations had been supported by the programs under evaluation. This review provided the team with participation status (to determine whether we were evaluating participant or nonparticipant SO) and program-attributable savings at the customer level.
- **Customer Self-Reports.** Telephone surveys of the LED owners in the on-site sample to collect information on the influence of the program on participant and nonparticipant customer. This survey effort provided customer-level FR and SO rates.

The analysts then used the standard SO and FR methodology from the 2013 Commercial and Industrial Electric Program Free-ridership and Spillover Study to assess the impact of the Pas' programs on LED purchases within Massachusetts, using the following formulas to estimate population level FR and SO. The analysts needed to estimate FR for this specific sample since FR was not estimated separately for LEDs in the 2013 Commercial and Industrial Electric Programs Free-ridership and Spillover Study.

Equation 1 through **Equation 3** show how free ridership and spillover rates were calculated. Any respondent to the survey that identified as a participant was asked the free ridership series of questions. In the example provided in Table B-3 below, the team included all self-identified participants in the calculation of the free-ridership rate. The final NTG recommended to the PAs did not include responses from nine unverified participants.

Equation 1: Free-Ridership Calculation

$$\frac{\sum(\boldsymbol{Free\ Rider\ Rate\ i} * \boldsymbol{Site\ Weight\ i} * \boldsymbol{Program\ LED\ Savings\ i})}{\sum(\boldsymbol{Site\ Weight\ i} * \boldsymbol{Program\ LED\ Savings\ i})}$$

*where i refers to an individual site

The team calculated participant spillover using responses, savings and site weights associated with program participants. Non-participant spillover was calculated using information associated with respondents who had received program-rebated LEDs. Although the survey asked self-identified participants a separate set of questions than non-participants, each set of questions contained the same basic spillover questions which allowed the team to analyse NTG by assigned unverified participants as participants or non-participants. In the results shown in Table B-3 unverified participants

are analysed as participants, thus included in participant spillover, whereas there results were not included in the final NTG ratio[69].

Equation 2: Participant Spillover Calculation

$$\frac{\sum(\textbf{\textit{Participant Spillover Rate } i * \textit{Site Weight } i * \textit{Out} - \textit{of} - \textit{Program LED Savings } i)}{\sum(\textbf{\textit{Site Weight } i * \textit{Out} - \textit{of} - \textit{Program LED Savings } i)}$$

where i refers to an individual site

Equation 3: Nonparticipant Spillover Calculation

$$\frac{\sum(\textbf{\textit{Non} - \textit{Participant Spillover Rate } i * \textit{Site Weight } i * \textit{Site} - \textit{level LED Savings } i)}{\textbf{\textit{Total } 2011 - 2014 \textit{ Program Savings}}}$$

where i refers to an individual site

Table B-3 summarizes the results of the computations shown above.

Table B-3. Summary of Total Savings v. Baseline, Free-Ridership, Spillover, and Net Savings

Net Savings Component	Net Savings Component LED Savings (Numerator for Equations 1-3)	Total Sampled LED Savings* (Denominator for Equations 1-3)	Free-Ridership or Spillover Rate
Free-Ridership	82,494 MWH	385,606 MWH**	21%
Participant Spillover	166,267 MWH	246,751 MWH	67%
Nonparticipant Spillover	6,778 MWH	542,972 MWH	1%

* for non-participant spillover this represents total 2011-2014 LED savings

** = Gross Savings in the Sample Facilities

Using the results shown above, the NTG ratio for LEDs was calculated as shown in **Equation 4**.

Equation 4: Net to Gross Calculation

NTG = 1 - Free-Ridership + Participant Spillover + Nonparticipant spillover

NTG = 1 – 0.21 + 0.67 + 0.01 = 1.47

We note that respondents were asked to verify their participation status at the beginning of the telephone survey; those identified as nonparticipants in the tracking data review that self-identified as a participant were ultimately not included in the FR or SO calculations due to this discrepancy. The evaluation team estimated the impact of including these respondents in participant and nonparticipant SO estimates and discussed the impacts of each scenario with the PAs and EEAC Consultants. The team decided to exclude these respondents from the analysis as a result of the uncertainty as to what portion of the LED installations in the non-verified sites should be allocated to tracked savings versus SO.

[69] Responses from the unverified participants could not be used in the calculation of free-ridership as they did not have LED savings recorded in the tracking data.

B.4 EXAMPLE 4: HIGH BAY LIGHTING MARKET EFFECTS

For this study of market effects of programs to promote energy-efficient high bay lighting, recent on-site surveys of population samples of commercial customers were not available. Moreover, commercial saturation surveys typically do not disaggregate conditioned space into high bay and standard ceiling heights. Thus, analysts needed to generate estimates of high bay floor space through primary research. Finally, at the time of the survey, at least five different high bay lighting technologies commanded substantial market share. Table B-4 illustrates how analysts used a wide variety of sources including primary customer telephone surveys, and surveys of samples of contractors in the program and comparison areas to estimate technology market share and gross savings. We then subtracted high bay program savings net of FR to estimate total out-of-program savings. We then used more qualitative methods to attribute portions of total out-of-program savings to net program savings and other programs.

Table B-4. Estimate of Demand and Energy Use Reductions: Massachusetts versus the Comparison Area

Row	Item	Value	Notes/Sources
1	Total square feet served by 2007-2010 HBL purchases in Massachusetts	467 million	Estimated from Massachusetts end-user survey
2	Average Watts per square foot (lighting power density): Program Area Efficacy	0.65 w/sf	Estimated based on technology share results from the Massachusetts contractor survey
3	Average watts per square foot (lighting power density): Baseline Efficacy	0.72 w/sf	Estimated based on technology share results from the Comparison Area contractor survey
4	Total MW of high bay lighting purchased: Program Area	304.7 MW	Row 2 * Row 1
5	Total MW of high bay lighting purchased: Baseline Efficacy	335.2 MW	Row 3 * Row 1
6	Difference in MW installed: Program Area v. Baseline Efficacy	30.5 MW	Row 5 – Row 4
7	Difference in GWh/Year Usage: Program Area v. Baseline Efficacy	90.7 GWh/Year	Row 6 * average 2,975 annual operating hours
8	Net savings estimated from 2010 impact evaluations	27.6 GWh/Year	Program-level evaluations
9	Savings from out-of-program adoptions, net of baseline adoptions	63.0 GWh/Year	Row 7 – Row 8

B.5 REFERENCES FOR RESIDENTIAL STUDIES NOTED IN TABLE A-4

2012

a. *Home Energy Services Net-to-Gross Evaluation.* June 2012 http://ma-eeac.org/wordpress/wp-content/uploads/Home-Energy-Services-Net-to-Gross-Impact-Evaluation_Part-of-the-Massachusetts-Residential-Retrofit-Low-Income-Program-Area-Evaluation.pdf
b. *Massachusetts Multifamily Program Impact Analysis.* July 2012. http://ma-eeac.org/wordpress/wp-content/uploads/Massachusetts-Multifamily-Program-Impact-Analysis-Report-Appendix.pdf
c. *Home Energy Services Impact Evaluation.* August 2012 http://ma-eeac.org/wordpress/wp-content/uploads/Home-Energy-Services-Impact-Evaluation-Report_Part-of-the-Massachusetts-2011-Residential-Retrofit-and-Low-Income-Program-Area-Evaluation.pdf
d. *Wi-Fi Programmable Controllable Thermostat Pilot Program Evaluation.* September 2012. http://ma-eeac.org/wordpress/wp-content/uploads/Wi-Fi-Programmable-Controllable-Thermostat-Pilot-Program-Evaluation_Part-of-the-Massachusetts-2011-Residential-Retrofit-Low-Income-Program-Area-Study.pdf

2013

a. *2012 Residential Heating, Water Heating, and Cooling Equipment Evaluation: Net-to-Gross, Market Effects, and Equipment Replacement Timing.* June 2013. http://ma-eeac.org/wordpress/wp-content/uploads/2012-Residential-Heating-Water-Heating-and-Cooling-Equipment-Evaluation_Net-to-Gross-Market-Effects-and-Equipment-Replacement-Timing-Volume-I-June-2013.pdf
b. *2012 Home Energy Services Pre-Weatherization Initiative Evaluation.* April 2013. http://ma-eeac.org/wordpress/wp-content/uploads/Year-2012-Home-Energy-Services-Pre-Weatherization-Initiative-Evaluation-Final-Report-April-2013.pdf
c. *Massachusetts Cross-Cutting Behavioral Program Evaluation Integrated Report.* June 2013. http://ma-eeac.org/wordpress/wp-content/uploads/Cross-Cutting-Behavioral-Program-Evaluation-Final-Integrated-Report-June-2013.pdf

2014

a. *Massachusetts Residential New Construction Net Impacts Report.* January 27, 2014. http://ma-eeac.org/wordpress/wp-content/uploads/Residential-New-Construction-Net-Impacts-Report-1-27-14.pdf
b. *Market Lift Assessment.* September 2014. http://ma-eeac.org/wordpress/wp-content/uploads/Market-Lift-Assessment-Final-Report1.pdf
c. *Recommended Methods for Assessing Market Effects of HVAC Programs.* November 25, 2014. http://ma-eeac.org/wordpress/wp-content/uploads/Recommended-Methods-for-Assessing-Market-Effects-of-HVAC-Programs.pdf

2015

a. *Massachusetts Cross-Cutting Behavioral Program Evaluation Opower Results Memo.* March 2015. http://ma-eeac.org/wordpress/wp-content/uploads/Behavior-Program-Impact-Evaluation-Memo.pdf
b. *Massachusetts Upstream Lighting Program Net-to-Gross Ratio Estimates Using Supplier Self-Report Methodology.* March 2015. http://ma-eeac.org/wordpress/wp-content/uploads/Upstream-Lighting-Net-to-Gross-Estimates-Using-Supplier-Self-Report-Methodology.pdf
c. *Preliminary Lighting Demand Elasticity Findings Memo.* March 16, 2015. http://ma-eeac.org/wordpress/wp-content/uploads/Preliminary-Lighting-Demand-Elasticity-Findings-Memo.pdf

d. *MA Residential Point-of-Sale Modeling.* February 2015. http://ma-eeac.org/wordpress/wp-content/uploads/Residential-Point-of-Sale-Modeling-Final-Report.pdf
e. *Multistage Lighting Net-to-Gross Assessment.* August 2015. http://ma-eeac.org/wordpress/wp-content/uploads/Multistage-Lighting-Net-to-Gross-Assessment-Overall-Report.pdf
f. *Recommended Methods for Assessing Market Effects of Non-residential New Construction Programs.* November 25, 2014. http://ma-eeac.org/wordpress/wp-content/uploads/Recommended-Methods-for-Assessing-Non-residential-New-Construction-Market-Effects.pdf
g. *Final Draft Report of Massachusetts LED Market Effects: Baseline*. March 1, 2015. Characterization http://ma-eeac.org/wordpress/wp-content/uploads/LED-Market-Effects-Baseline-Characterization-Final-Draft.pdf
h. *Methods for Measuring Market Effects of Massachusetts Energy Efficiency Programs*. November 14, 2015.
i. *Top-down Modeling Methods Study.* March 31, 2015. http://ma-eeac.org/wordpress/wp-content/uploads/Top-down-Modeling-Methods-Study-Final-Report.pdf

2016

a. *Berkshire Gas Home Energy Report Program Evaluation.* January 25, 2016. http://ma-eeac.org/wordpress/wp-content/uploads/Berkshire-Home-Energy-Report-Process-and-Impact-Evaluation-Final-Draft-2016-01-25-1.pdf

B.6 REFERENCES FOR C&I STUDIES NOTED IN TABLE A-5

a. *HBL Market Effects Study*. June 17, 2011. http://ma-eeac.org/wordpress/wp-content/uploads/High-Bay-LIghting-Market-Effects-Study-Final-Report.pdf
b. *LED Market Effects: Baseline Characterization.* March 1, 2015. http://ma-eeac.org/wordpress/wp-content/uploads/LED-Market-Effects-Baseline-Characterization-Final-Draft.pdf
c. *Massachusetts LED Spillover Analysis.* September 24, 2015. http://ma-eeac.org/wordpress/wp-content/uploads/MA_LED_Spillover_FINAL-Report_092415.pdf
d. *Commercial and Industrial Lighting and Controls Market Effects Study.* In Progress.
e. *Process Evaluation of the 2012 Bright Opportunities Program.* June 14, 2013. http://ma-eeac.org/wordpress/wp-content/uploads/Process-Evaluation-of-the-2012-Bright-Opportunities-Program-Final-Report-6.14.13.pdf
f. *Code Compliance Baseline Study.* August 24, 2012. *http://ma-eeac.org/wordpress/wp-content/uploads/Code-Compliance-Baseline-Study.pdf*
g. *Top-down Modeling Methods Study.* March 31, 2015. http://ma-eeac.org/wordpress/wp-content/uploads/Top-down-Modeling-Methods-Study-Final-Report.pdf
h. *2014-2015 Commercial and Industrial Natural Gas Programs Free-ridership and Spillover Study.* August 10, 2015. http://ma-eeac.org/wordpress/wp-content/uploads/CI-Natural-Gas-Programs-Free-ridership-and-Spillover-Study.pdf
i. *2013 Commercial and Industrial Electric Programs Free-ridership and Spillover Study.* February 17, 2015. http://ma-eeac.org/wordpress/wp-content/uploads/CI-Electric-Programs-Free-Ridership-and-Spillover-Study.pdf
j. *Upstream HVAC Market Penetration Data Collection.* In Progress.
k. *Boiler Market Assessment Phase II.* In Progress.
l. *Assessment of the Share of Incentivized High Efficiency Equipment.* Planned.

APPENDIX C: REFERENCES

Arkansas Public Service Commission, (2016). *Arkansas Technical Reference Manual, Version 6.0.* http://www.apscservices.info/EEInfo/TRM6.pdf.

The Cadmus Group and Navigant, (2012). *Home Energy Services Net-to-Gross Evaluation*.

DNV-GL, (2016). *Massachusetts Commercial/Industrial Baseline Framework*.

DNV GL, NMR Group, Inc. and Tetra Tech, (2015). *Recommended Methods for Assessing Market Effects of Non-residential New Construction Programs*.

DNV GL, (2016). *Massachusetts C&I Market Characterization On-Site Assessments and Market Share and Sales Trends Study*. http://ma-eeac.org/wordpress/wp-content/uploads/MA-CI-Market-Characterization-Study.pdf.

Eto, J.; Prahl, R.; Schlegal J., (1996). *A Scoping Study on Energy-efficiency Market Transformation by California Utility DSM Programs*. http://emp.lbl.gov/sites/all/files/lbnl%20-%2039058.pdf.

Illinois Energy Efficiency Stakeholder Advisory Group, (2016*). Illinois Statewide Technical Reference Manual – Attachment A: IL-NTG Methodologies*.

Tetra Tech, KEMA, NMR Group, Inc., (2016). *Net-to-Gross Methodology Research Stage 3 Plan—TXC 8*.

Illinois Energy Efficiency Stakeholder Advisory Group, (2016). *Illinois Statewide Technical Reference Manual for Energy Efficiency, Version 5.0. Volume 4: Cross-Cutting Measures and Attachments.*

Johnson Consulting Group, (2014). *Decision-Framework for Determining Net Savings Approach—Supplemental Document #2 to Principles and Guidance*.

KEMA, (2011). *HBL Market Effects Study; Project 1A New Construction; Market Characterization*,. http://ma-eeac.org/wordpress/wp-content/uploads/High-Bay-LIghting-Market-Effects-Study-Final-Report.pdf.

Navigant and Illume, (2015). *Massachusetts Cross-Cutting Behavioral Program Evaluation OPower Results.*

NMR Group, Inc. and Tetra Tech, (2014). *Recommended Methods for Assessing Market Effects of HVAC Programs.*

NMR Group, Inc., (2014). *Residential New Construction Net Impacts Report.*

NMR Group, Inc., Cadmus, and DNV GL, (2015). *Multistage Lighting Net-to-Gross Assessment: Overall Report.*

NMR Group, Inc. and Tetra Tech, (2015). *Recommended Methods for Assessing Market Effects of CI Lighting and Controls Programs.*

NMR Group, Inc., EcoMetric Consulting, Demand Side Analytics, (2016). *Evaluation Framework for Pennsylvania Act 129 Phase III Energy Efficiency and Conservation Programs*

Prahl, et al., (2008). *Integrating Supply-Side Results With End-User Net-to-Gross Self Reports.*

State of Iowa, Department of Commerce, (2015). *Iowa Energy-Efficiency Net-to-Gross Report.*

Swann, William B., Jr., (2011). *Self-Verification Theory.* [In P. Van Lange, A.W. Kruglanski, and E.T. Higgins (eds.), Handbook of Theories of Social Psychology].

Tetra Tech, KEMA, NMR Group, Inc., (2011). *Cross-Cutting (C&I) Free-Ridership and Spillover Methodology Study Final Report.*

Tetra Tech, NMR Group, Inc., KEMA, (2011). *Cross-Cutting Net to Gross Methodology Study for Residential Programs – Suggested Approaches.*

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