Northeast Energy Efficiency Partnerships

Current and Evolving Policies, Issues, and Methods Pertaining to Gross and Net Savings – *Supplemental Document #1 to Principles and Guidance*

April 2016

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

REGIONAL EVALUATION,
MEASUREMENT & VERIFICATION FORUM

About NEEP

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

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

About Navigant Consulting

Navigant is a leading Demand Side Management consultant company in smart grid and energy efficiency research, planning and evaluation, providing services to program administrators across North America.

About Tetra Tech

Founded in 1966, Tetra Tech is a leading provider of consulting, engineering, program management, construction, and technical services addressing the resource management and infrastructure markets. Staff in Tetra Tech's Madison office have been directing evaluation, measurement and verification (EM&V) studies for electric and natural gas utilities throughout the United States for over two decades.

Acknowledgements

This document was prepared by:
Daniel Violette, Navigant Consulting
Teresa Lutz and Pam Rathbun, Tetra Tech

With project management by NEEP staff:
Elizabeth Titus and Julie Michals

With contributions by:
Richard Ridge, Ridge & Associate
Miriam Goldberg and Noel Stevens, DNV GL
Brian Buckley, NEEP

This project was informed by the project subcommittee with representation from the participating states. Financial support or involvement with this project and development of these guidelines does not necessarily indicate full agreement with all contents within nor does it bind participants to its content or preclude them from taking future positions in public proceedings or other venues that may vary from these guidelines.

This project also benefited from input from the Forum's subcommittee and by industry expert review, both inside and outside of NEEP. These efforts are greatly appreciated and provided new insights and expanded thinking about the challenges and issues related to gross and net savings estimation. However, the project team bears the responsibility for the contents of the report.

Table of Contents

Preface

This document - Current and Evolving Policies, Issues, and Methods Pertaining to Gross and Net Savings - was prepared for the Regional Evaluation, Measurement, and Verification Forum ('the Forum'). This document supplements the Principles and Guidance document by adding information on evolving policies, some selected issues, and some trends in energy efficiency evaluation methods. The information in this document was selected to provide additional context and emerging examples of how gross and net savings can be measured or applied.

The Forum, established in 2008, is a regional project facilitated and managed by Northeast Energy Efficiency Partnerships (NEEP). The Forum supports the development and use of consistent and transparent protocols to evaluate, measure, verify, and report the savings, costs, and emissions impacts of energy efficiency and other demand-side resources. The Guidance emerged out of recognition among Forum member states of the importance of understanding how states define, estimate, and apply gross and net savings across the region.

> *The goal of this effort is to provide a framework that can be used or adapted to help stakeholders assess gross savings and net savings issues in the context of overall EE policy decisions.*

Taken together, the Gross and Net Savings: Principles and Guidance, this document, and Supplemental Document #2: Decision-Framework for Determining Net Savings are meant to provide a set of principles to inform states' decisions regarding their applications of gross and net savings based on policy goals. The guidance and supplemental documents do not present arguments for specific regulatory policies nor do they advocate for the use of specific evaluation methods. The goal of this effort is to provide a framework that can be used or adapted to help stakeholders assess gross savings (GS) and net savings (NS) issues in the context of overall energy efficiency (EE) policy decisions.[1]

EE policy is in the midst of dynamic evolution, as utilities are transitioning to a 'next generation' in which EE becomes more integrated with other distributed resources and where technologies and advanced analytical software make customer data more available These principles and guidance is designed to be useful in assessing policy decisions in this changing industry context.

[1] There are many nuances that will come up in any GS and NS policy discussion. It is not possible to anticipate and address each of these within this guidance paper. The authors are keenly aware of many additional points and concepts that could have been addressed. This guidance document attempts to establish an appropriate compromise by providing a useful overall framework without becoming a technical paper.

1. Introduction

This document is a supplement to the "Gross Savings and Net Savings: Principles and Guidance" main report. This document is meant to augments the principles and guidance document by presenting information on issues that may influence EE evaluation and the way gross savings (GS) and net savings (NS) are being used in assessing EE accomplishments. Three sections are presented. The first presents trends in the way (NS) is used in policy frameworks across jurisdictions in the United States. The second section presents potential issues and influences of two policy discussions: 1) the EPA's rulemaking on Clean Power Plans (CPP), and the related draft of EPA's guidance on EM&V for EE in these plans; and 2) New York's work on Reforming the Energy Vision (REV). These are examples of two policy discussions that are influencing EE policy.

Each of these sections is meant to provide context and highlight issues that might influence *future* EE policies, and emphasize the need to periodically review policy decisions related to NS to ensure they remain relevant.

2. Trends in EM&V Methods in North America and the Northeast

EM&V policies and methods are reviewed by at least some jurisdictions every year. A change in EE targets, cost recovery, and/or incentives can also trigger a review of the EM&V methods and requirements. Based on recent surveys, a relationship exists between certain EE policies adopted by a state and EM&V requirements, as well as how NS is defined and treated.

This section uses the results of a survey[2] conducted in the first half of 2015 to show trends in NS methods and the way the estimates are used by states. The American Council for an Energy-Efficient Economy (ACEEE's) 2014 report was a key resource for investigating the state of policies in various jurisdictions around the country.[3] States, regulators, and policymakers bring diverse NS perspectives to the table. A number of states focus on GS for program reporting (e.g., Pennsylvania), but encourage (and sometimes require) NS or NTG ratio[4] estimates to help improve program design. In fact, no state was found that would prohibit NS research for improving program design. To ensure a common platform for comparison, the reported results focus on two elements: 1) NS policy with regards to the reporting of energy savings, and 2) the use of NS as measured against EE goals.

[2] See Violette et al.,(2015). Other links to this work including the Commission Order and comments by parties can be found at: Commission Order:
https://efs.iowa.gov/cs/groups/external/documents/docket/mdax/mziy/~edisp/1322686.pdf
Comments by Parties:
https://efs.iowa.gov/cs/groups/external/documents/docket/mdax/ndy3/~edisp/1467318.pdf
https://efs.iowa.gov/cs/groups/external/documents/docket/mdax/ndy3/~edisp/1467085.pdf

[3] Examining the Net Savings Issue: A National Survey of State Policies and Practices in the Evaluation of Ratepayer-Funded Energy Efficiency Programs. American Council for an Energy-Efficient Economy (ACEEE), 2014

[4] The net-to-gross (NTG) ratio is the net savings estimate divided by the gross savings estimate.

- Overall NS Policy. This shows whether or not program administrators must report savings and assessment against goals at the gross or the net level. Note that states that assume a NTG of 1.0 are assumed to effectively be gross states since there are no upward or downward adjustments due to program attribution.

- Definition of NS (Allowance for Spillover). Within the NS jurisdictions, there is wide variation in which aspects of net savings are allowed in terms of savings claims. Some states consider net of free ridership—not counting any aspects of spillover—to be NS. Other states allow different aspects of spillover (i.e., participant and non-participant) to be counted as achieved NS.

- NS Methods Protocols. Certain states (such as California and Massachusetts) have developed NS method protocols that recommend specific approaches, and in some cases, specifics regarding the calculations (e.g., survey batteries and analysis algorithms for self-report approaches). The use of these may be recommended, or in certain states (e.g., Pennsylvania), an emerging practice is to require the use of the methods protocols.

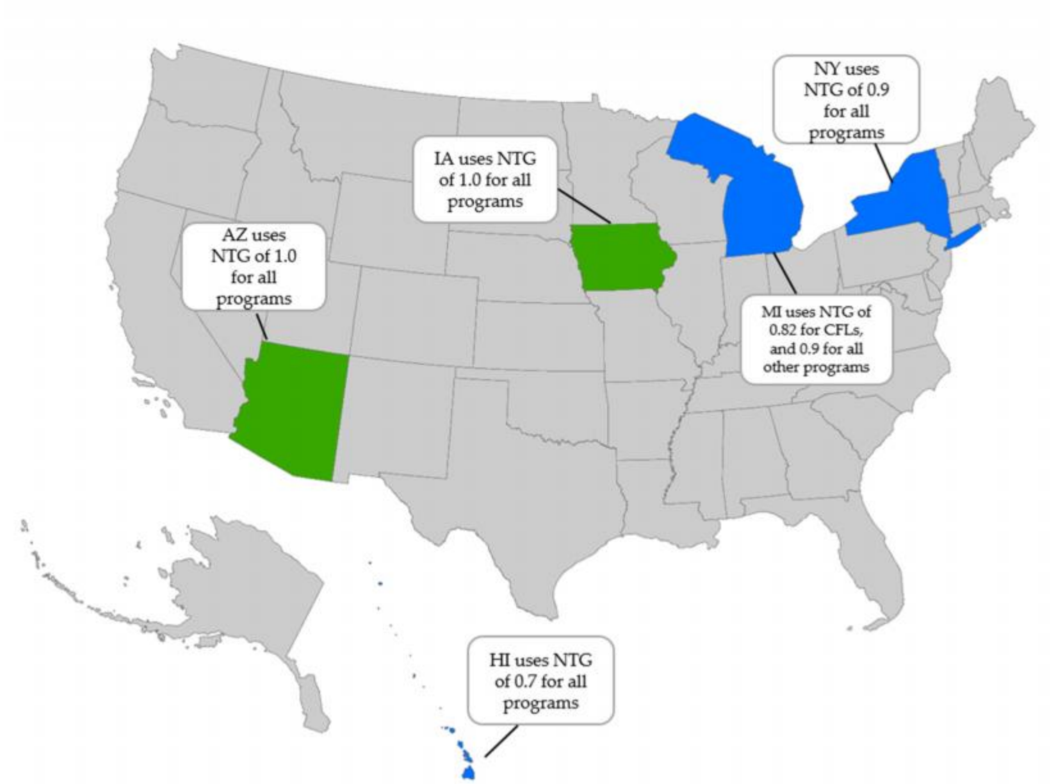
- Fixed or Researched NS. A number of NS states lock in a fixed NS value that applies to all, or at least most, programs. Note that while this has a prospective aspect to it (in that NS is fixed prior to the program year), this is considered different than our definition of prospective NS (below), which is typically based on researched values that can vary by program and measure

- Prospective versus Retrospective Application of NS Values. Another emerging practice is the move toward the prospective use of NS values, whereby NS values researched in a current program year are applied prospectively to future year(s), rather than retrospectively to the current or past program year(s). Once NS values are established, they are essentially locked until an updated value is derived and applied prospectively.

2.1 Overall NS Policy: Gross versus Net

While some states adopt a net policy and others adopt a gross policy, a number of net states deem all program NTG values at 1.0 or a different value. As shown in Figure 2-1, several states have a policy 1.0 for all programs, while other states deem a value other than 1.0 for all programs. For example, Michigan deems all values at 0.9, with the exception of CFLs which have their own deemed NTG value.

Figure 2-1. Net States with Deemed NTG Values



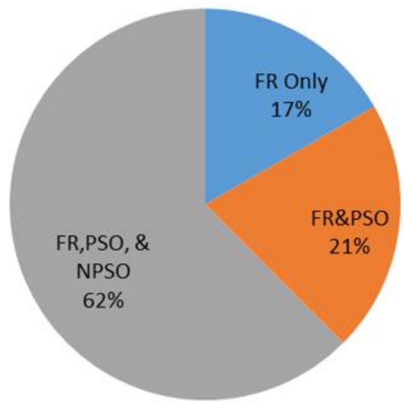
Source: Violette et al. (2015) and Navigant research
(Note: This research was completed in March, 2015. Some circumstances may have changed in a few states. In particular, the NTG policies New York is are in the process of revision)

2.2 Definition of NS (Allowance for Spillover)

As shown in Figure 2-2, nearly two-thirds (62%) of those jurisdictions that use NS allow for free ridership, participant spillover, and non-participant spillover, while 21% allow for free ridership and participant spillover but do not allow for non-participant spillover.[5] Only 17% of the jurisdictions with NS (a total of four states) limit NS to net of free ridership (i.e., do not allow for the consideration of any contributions from spillover to count toward the net savings estimates).

[5] Note that in this survey, market effects is included as a subset of non-participant spillover rather than breaking it out separately, because we are not considering it in the context of market transformation studies, but rather as a subcategory of non-participant spillover. Precedent for this distinction is set in the U.S. Environmental Protection Agency (EPA) in the Model Energy Efficiency Program Impact Evaluation Guide: A resource of the National Action Plan for Energy Efficiency (NAPEE), available at: http://www.epa.gov/cleanenergy/documents/suca/evaluation_guide.pdf

Figure 2-2. Inclusion of Free Ridership (FR), Participant Spillover (PSO), and Non-Participant Spillover (NPSO) in NTG Ratios (n=24)



Source: Navigant Team research

2.3 Fixed or Researched NS

Three states—Hawaii, Michigan, and New York—rely on fixed NTG values that differ from 1.0 for all programs within the energy-efficiency portfolios. As shown in Table 2-1, the ratios range from 0.7 to 0.9, and Michigan lowered the NTG for CFLs based on research showing a lower NTG than the other measures in the portfolio. Arkansas used a similar approach during the first year of program implementation in 2011 by having all programs use a stipulated NTG of 0.8. CFLs were the exception, and were required to use a NTG of 0.62. After the first year, the programs were required to rely on researched values.

Using fixed, portfolio level NTG values is most reflective of the deemed view perspective, implicitly implying that having distinct NTG values for each program and/or measure are likely too imprecise, and that overarching adjustments – in some cases based on findings from other jurisdictions – are adequate.

Table 2-1. Values for Fixed NTG Ratios

State	Fixed NTG Ratio
Hawaii	0.7
Michigan	0.9 (0.82 for CFLs)
New York	0.9 (under revision in NY)

Source: Violette et al. 2015 and Navigant Team research

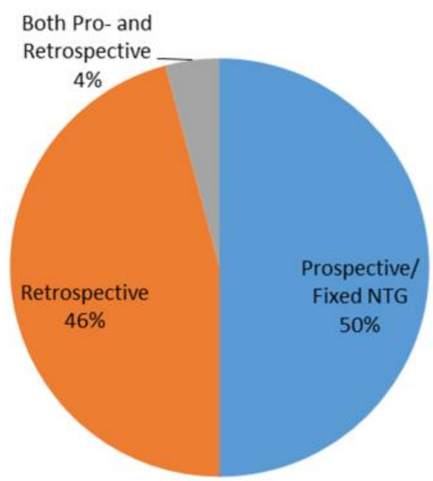
2.4 Prospective versus Retrospective Applications

Because NTG ratios can be so volatile and the elasticity of the findings is 1.0 (i.e., a 1% drop in NTG is a 1% drop in attributable savings), program administrators have perceived significant risk and uncertainty with retrospective NTG application. Thus, many states have now moved to using NTG results prospectively, rather than retrospectively. As described above, prospective NTG means that any updates to NTG values are applied in future program years, not in the year in which they are developed or to prior program years. As shown in Figure 2-3, half of the jurisdictions with NS (50%) use either a fully prospective/fixed NTG or a combination of prospective and retrospective.[6]

While this method can significantly reduce risk for program administrators, to be effectively implemented it requires careful planning in terms of the timing and nature of the NTG research so that the results are applicable going forward. Research needs to be updated as markets and incentive structures and program design change, and secondary research from other similar programs may be helpful in determining NTG ratios.

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 actual occurring. 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), as well as the deemed view (i.e., NTG has significant uncertainty so it is essential to lock down values prior to program implementation).

Figure 2-3. Use of Prospective vs. Retrospective NTG (n=24)



Source: Navigant Team research

[6] California, as part of the Efficiency Savings and Performance Initiative (ESPI), allows prospective NTG for some measures but retrospective NTG for measures that are determined to be less stable in terms of program attribution.

2.5 Relationship between NS Policy and Other DSM Policy Objectives

It is important to understand NS policies in the context of other DSM policies in a jurisdiction. Certain jurisdictions may have DSM policies—particularly financial policies—where the importance of more precise, fully attributable savings estimations may be perceived as more important. In addition, other DSM policies may be related to the treatment of NS. The survey examined a number of DSM policies and their relationships to how NS estimates are applied:

1. Energy Efficiency Resource Standards (EERS): State-level policy that sets long-term mandatory energy savings targets for utilities and energy-efficiency program administrators.

2. Decoupling: A regulatory tool that serves as a means of helping utilities overcome the throughput incentive; i.e., the contribution to gross income that occurs with every energy unit sold because the unit (variable) price recovers some of a utility's fixed costs. A decoupling mechanism separates a utility's revenue from its unit sales volume without affecting the design of customer rates.

3. Lost Revenue Recovery: Allows a utility to recover the lost revenue attributable to DSM programs by increasing revenue by that same amount. Can be based on decoupling (see above) or by adjustments (rate adjustment).

4. Risk-Reward Mechanisms: Allows utilities to earn bonuses for meeting or exceeding goals, or imposes financial penalties for savings shortfalls.

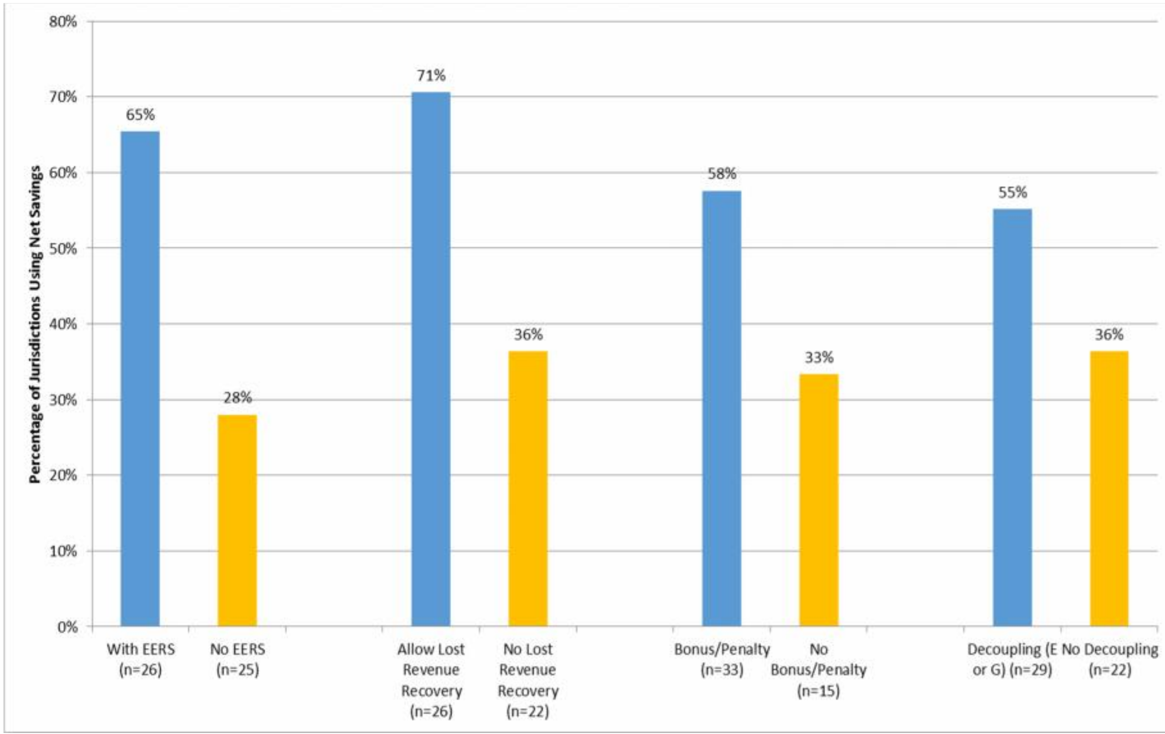
The charts below focus on trends and correlations between NS policies and the broader DSM policy objectives listed above. Note that the sample sizes are small, and each subset of analysis (e.g., examining only states that have NS policies) leads to even smaller sample sizes. The results, therefore, need to be used with caution; however, there are some strong correlations suggestive of a relationship between the DSM policies listed above and treatment of NS.

The analysis focused on two important NTG policies: 1) whether savings at a gross or a net level are required; and 2) for those jurisdictions that do use NS, whether or not the NTG is applied prospectively or retrospectively. These NTG policies are then compared against the different DSM policies outlined above.

As shown in Figure 2-4, jurisdictions that have an EERS, allow lost revenue recovery, and have bonuses or penalties tend to also require net, rather than gross, savings. These differences are substantial; for example, 71% of the jurisdictions that allow for lost revenue recovery also require NS, while only 36% of the jurisdictions that do not allow for lost revenue recovery require NS. Each of these DSM policies with greater correlation with NS also tend to be associated with increased stakes—particularly financial—in terms of the outcome. In other words, states where there are potentially millions of dollars at stake on the outcome of the savings assessment also tend to use net, rather than gross, savings.

Interestingly, decoupling is also associated with NS. This may reflect program administrators' interest in mitigating risk due to DSM—and possibly NS—through the use of decoupled rates.

Figure 2-4. Percentage of Jurisdictions with Net Savings that also have a DSM Policy

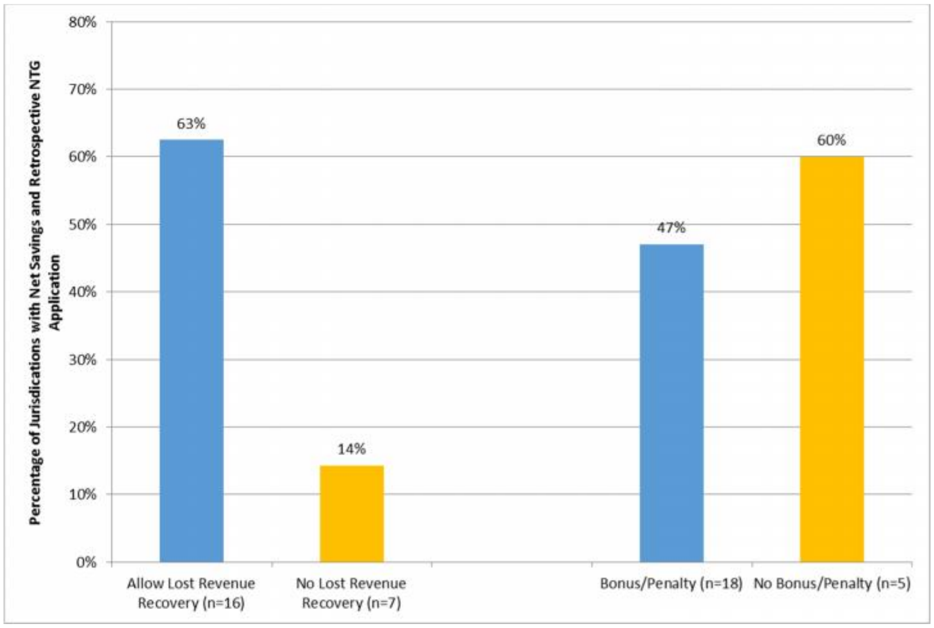


(For example, the first blue bar indicates that 65% of the states that have an EERS require net savings and the second blue bar says that 71% of the states that allow for lost revenue recovery also require net savings.)

Source: Navigant Team research

The survey also examined the relationship of prospective versus retrospective NTG application and other DSM policy objectives. Due to the small sample size, the analysis was limited to two DSM financial policies: lost revenue recovery and risk-reward mechanisms. As shown in Figure 2-5, jurisdictions with lost revenue recovery are much more likely to have retrospective application of NTG findings versus areas without lost revenue recovery (63% vs. 14%). The relationship of risk-reward mechanism and the use of prospective versus retrospective NTG application was less clear, particularly since only five states with NS did not have a bonus or penalty.

Figure 2-5. Percentage of Jurisdictions with Retrospective Net Savings (or NTG) Application by Other DSM Policy (States with Net Savings Only)[7]



Source: Navigant research

These results show that, in general, decoupling, lost revenue recovery, or risk-reward mechanisms tend to lessen the financial stakes in the outcome of the savings findings, and tend to be associated with the use of gross, rather than net, savings. In addition, those states that do not have lost revenue recovery but still require NS tend not to require retrospective NTG application (i.e., they tend to use prospective NTG).

2.6 Summary – Trends in Net Savings Policies

There are several conclusions that can be drawn from the recent survey efforts on NS policies:

- There has been a steady increase in the number of jurisdictions that now allow for the inclusion of spillover in the estimates of NS. Nearly two-thirds (62%) of those jurisdictions nationally that use NS allow for the inclusion of FR, PSO, and NPSO factors. Another 21% allow for FR and PSO, but do not allow for NPSO.[8] Only 17% of the jurisdictions only defined NS as being net of FR.
- A second trend is an increasing move towards jurisdictions using NS (or NTG ratios) prospectively, rather than retrospectively. The reason given for this is that NS values and NTG ratios applied retrospectively are perceived by program administrators as posing significant

[7] Note that since California has both prospective and retrospective NTG it is not included in this chart.

[8] On anecdotal reason for this expressed by some jurisdictions was that estimates of NPSO are judged as having less certainty in the magnitude of the estimates, even if there was some that viewed NPSO as existing.

risk and uncertainty with respect to planned targets and financial incentives that were set based on approved programs.

- NS plays a larger role if there are other EE policies including EERS, lost margins, decoupling, or risk reward mechanisms. The absence of these policies in a jurisdiction tends to lessen the stakes in the outcome of the savings findings.

3. Evolving Policies

This section discusses two evolving policies: 1) The Clean Power Plan and the EM&V Guidance provided by EPA last fall, and 2) the New York Reforming the Energy Vision (NY REV). States besides New York such as California and Minnesota are also looking at various electricity reforms, and there are other important policy discussions looking at distributed energy resources that have the potential to change the electricity landscape. The two evolving policies selected for this section represent two areas that may be driving evolution in EE as well.

3.1 NY REV and other states looking at distribution system platforms

The discussions taking place in a number of jurisdictions on the appropriate role of the regulated utility may also influence EE evaluation and the role of NS. California, Minnesota,[9] and New York are leaders in a discussion about alternative business models for distribution utilities. These new business models, in general, change the role of the distribution utility to be more of a services provider, e.g., a distribution system platform that promotes the appropriate adoption of distributed energy resources (DER) and continues its role in distributing electricity from wholesale sellers of electricity. This approach is meant to promote long-term economic resource allocation and a resilient electric system for customers.

Just like the restructuring of the electric system in the late 1990's, this change in business model may impact both EE programs and the way they are evaluated. From the New York State Public Service Commission (PSC), "REV is an energy modernization initiative that will fundamentally transform the way electricity is distributed and used in New York State."[10] EE is a distributed energy resource and it may need to be evaluated in terms of its synergies with other DER (e.g., distributed generation, storage, renewables, and demand response). EE can be an influence at the grid edge, most directly by reducing permanent peak loads in selected areas that can make the overall integration of DER more efficient. It is not clear how this might play out for EE, but it is an issue on the horizon.

A number of jurisdictions are pursuing this concept. NY REV is one of the leading efforts. In the past, New York utility efficiency programs encouraged individual customers to use more efficient equipment

[9] The debate on utility business models and the role of the utility in Minnesota is largely led by the Great Plains Institute's e21 initiative and portions of this approach have been adopted into legislation. See: http://www.betterenergy.org/projects/e21-initiative ; and http://www.utilitydive.com/news/how-the-e21-initiative-is-building-smarter-utility-business-models-in-minne/400781/ for links to this work.

[10] See: Reforming the Energy Vision Case 14-M-0101 Proceeding on Motion of the Commission in Regard to Reforming the Energy Vision Fact Sheet at http://www3.dps.ny.gov/.

and systems through direct rebates and subsidies. Under this approach, energy savings were acquired as a resource. This resource acquisition approach to efficiency can be contrasted to a market transformation approach, where the benefits of the program are defined in terms of wide-scale penetration and market acceptance of efficiency measures.

In February 2015, the New York Public Service Commission (PSC) released its final decision in Phase One of its Reforming the Energy Vision (REV) docket (NY PSC 2015a). According to the February 26, 2015 order, “REV will establish markets so that customers and third parties can be active participants, to achieve dynamic load management on a system-wide scale, resulting in a more efficient and secure electric system including better utilization of bulk generation and transmission resources. As a result of this market animation, distributed energy resources will become integral tools in the planning, management and operation of the electric system. The system values of distributed resources will be monetized in a market, placing DER on a competitive par with centralized options. Customers, by exercising choices within an improved electricity pricing structure and vibrant market, will create new value opportunities and at the same time drive system efficiencies and help to create a more cost-effective and secure integrated grid” (NY PSC 2015a, p. 11).

The focus of REV is on how energy is generated and consumed. The core objectives of REV include:

1. Enhancement of customer knowledge and tools to manage costs and energy consumption
2. Market animation and leverage of customer contributions
3. System wide efficiency, specifically addressing peak reduction utilizing a targeted methodology
4. Fuel and resource diversity
5. System reliability and resiliency
6. Reduction of carbon emissions.

Some of the drivers behind REV include intensive electric grid system upgrades between 2000 and 2010 and the ensuing consequences of Hurricanes Irene and Sandy in 2011 and 2012. In the REV policy framework and implementation plan, these trends are broken into four categories (NY PSC 2015a):

1. Regulatory models and economic efficiency – Challenges to traditional regulatory models include current system design, development and utilization, aging infrastructure, flat sales, fuel diversity, and the limitations of the system benefit charge. Opportunities for improvement include modernization of regulation and markets, infrastructure investment, stabilization of customer bills, optimization of fuel diversity, and realization of the potential of storage and innovation technologies.
2. System modernization for a digital economy - Challenges include information technology integration, cyber security. Opportunities include customer choice and animating markets, reliability and power quality, resilience, and system security.
3. Clean energy and environmental responsibility - Challenges include climate change, plug-in electric vehicles, combined heat and power, and integrating distributed renewable generation. Opportunities include reduced emissions and system heat rate, energy efficiency,

accommodating low-carbon generation, electrification of transportation systems, geothermal heating systems.

4. Universal service - Challenges include affordability, contraction of utilities' customer base. Opportunities include maintaining universal affordable service, securing utilities' financial stability.

As a result of state review, the REV initiative was instituted and is currently undergoing implementation. According to the latest Staff recommendations (NY PSC 2015b), an initial five-year Distributed System Implementation Plan is due from each utility on June 30, 2016, with reports due annually thereafter. In addition, utilities will work together to file a Supplemental DSOP by September 1, 2016, which will outline joint processes, protocols, and standards necessary for maintaining a modern grid and efficiently managing resources.

The emphasis of REV is on distributed energy sources and on shifting loads away from times of peak demand; however, a section of the decision (NY PSC 2015a, p. 72-82) specifically addresses EE. Under the decision, the utilities' post-2016 portfolio of EE programs will transition away from a focus on resource acquisition to a focus on long-term market transformation. The goal is to achieve greater market-wide efficiency savings with less direct ratepayer support and allow utilities more flexibility in achieving their goals cost-effectively as part of their ratemaking practice. The expectation is that market transformation programs will allow utilities to leverage more investment and fund more diverse and cost-effective measures than traditional subsidized rebate programs with fixed budgets.

Utilities are directed to submit their Energy Efficiency Transition Implementation Plans outlining energy efficiency programs they intend to implement in 2016 with the intention that they may be modified or supplanted by their Distributed System Implementation Plan (DSIP) as REV continues to develop. Utilities are still expected to maintain a portfolio of energy efficiency programs that meet or exceed current savings goals, but utilities will shift away from the resource acquisition model to the market transformation model utilizing program design such as whole-building management and demand response measures.

The final decision directs utilities to develop metrics for measuring the success of market transformation programs in consultation with NYSERDA and DPS staff. EM&V efforts should complement and not duplicate NYSERDA's evaluation efforts. REV is expected to result in more technology specific evaluation research and an added emphasis on M&V (e.g., project and technology-based reviews, TRM-based insights using in-situ metering, better defined load curves, market penetration rates). Advances in technology are expected to improve EM&V including evaluation guidelines, as well as data tracking and reporting, which will be continually revised to better address market transformation. With this in mind, program-based impact evaluations will be limited and will likely not address attribution or NS. Demand and energy savings calculations may be useful in evaluating short-term program effectiveness, but market transformation indicators are to be considered the true measure of program success. Market impacts will be measured through the 'test, measure, adjust' mechanism to provide quick cycle feedback, as well as through market characterization and market progress studies. These mechanisms will receive the largest share of funding. NYSERDA may be working regionally to measure program impacts in New York in comparison with other states and utilizing state-wide macro level accounting and potential studies.

3.2 EPA 111(d) and EPA EM&V Guidance

This section discusses the ongoing EPA rulemaking on the "Clean Power Plan" and the EPA Guidance on EM&V (EPA, 2015b) for estimating reductions in emissions from EE. Comments on the EPA EM&V Guidance were due on January 21, 2016,[11,12] but EPA has not yet responded to comments. In the interim, this section sets out some of the issues and ways in which the finalized rule and the related EPA Draft EM&V guidance relate to NS issues. In some ways, the finalized EPA guidance may have a reach beyond evaluation of EE for calculating emissions reductions for the Clean Power Plan. For example, the guidance may be applicable to other state air quality planning. The language and approaches in the rule and the guidance may be taken by some as a recommended approach for more broadly approaching energy savings estimation and use. It is believed that EPA's Draft EM&V Guidance will evolve due the comments already received, likely through additional comments or workshops.

As background, on June 18, 2014 the Environmental Protection Agency (EPA) issued a proposed rule entitled Carbon Pollution Emission Guidelines for Existing Stationary Sources: Electric Utility Generating Units (EPA 2014a) (more commonly referred to as the Clean Power Plan). This rulemaking, issued under the authority of the Clean Air Act (CAA) section 111(d), proposed emission guidelines for states to follow in developing plans to address greenhouse gas emissions from existing fossil fuel-fired electric generating units. On August 3, 2015, the EPA finalized the Clean Power Plan (CPP) rule to cut carbon pollution from existing power plants (EPA 2015a). The final rule was published in the Federal Register on October 23, 2015.

For goal setting purposes, the CPP sets carbon dioxide emissions performance rates for affected power plants that reflect the "best system of emission reduction" (BSER). EPA identified three "Building Blocks" as BSER:

1. Improved power plant efficiency
2. Shifting generation from fossil fuel electric generating units (EGUs) to natural gas power plants
3. Shifting generation to clean energy renewables, with emissions reductions to be accounted for in short tons of $C0_2$.

Energy efficiency was removed as a building block from the final rule. However, the final rule acknowledges that EE is an important and cost-effective way for states to lower CO_2 emissions in order to comply with CPP goals (EPA 2015c).

[11] NEEP joined several parties to file joint comments on the CPP and the EM&V Guidance. This can be found at the NEEP website at: http://www.neep.org/joint-ee-stakeholder-comments-epa-emv-clean-power-plan

[12] For more information on EPA's draft guidelines and the comment period see: https://www.epa.gov/cleanpowerplantoolbox/draft-evaluation-measurement-and-verification-guidance-demand-side-energy

Emission performance rates have been translated into equivalent state goals. States are required to develop plans to reach their goals starting in 2022. In order to maximize the range of choices available to states, EPA is providing state goals in three forms:

1. Rate-based goal measured in pounds of CO_2 per megawatt hour (lb/MWh)
2. Mass-based goal measured in short tons of CO_2
3. Mass-based goal with a new source complement.

In developing the state plan, a state has the flexibility to choose the measures that it prefers in order to achieve the CO_2 goals, including working with other states and emissions trading. States that choose to use energy efficiency as a strategy, can take two approaches: 1) a mass-based approach, where energy efficiency is automatically figured into emissions calculations and EM&V is generally not required; or 2) a rate-based approach, where the CPP allows verifiable energy savings from projects installed after 2012 to be applied toward emission rate credits (ERCs).

On August 3, 2015, the EPA also released two draft documents: 1) a federal plan for the Clean Power Plan that will also serve as a model rule for states developing their state plans,[13] and 2) the draft for public comment of its Evaluation, Measurement, and Verification (EM&V) Guidance for Demand-Side Energy Efficiency.[14] This EM&V guidance provides information to help states and energy efficiency providers implement the EM&V provisions, specifically for rate-based state plans in which compliance is measured. For mass-based state plans in which compliance is demonstrated by measuring emissions at the power plant, this guidance does not apply.

The rule itself addresses NS concepts related to EE evaluation. The final CPP rule discusses the EM&V plans and the procedures to quantify adjustments to emissions rates under the rule. The rule discusses key assumptions in the EM&V including the baselines from which the savings are to the calculated:

> *"A baseline that represents what would have happened in the absence of the EE intervention, such as the equipment that would most likely have been installed –or that a typical consumer or building owner would have continued using –in a given circumstance at the time of EE implementation." (CPP Final Rule, p. 1288)*

This statement by the EPA in the final rule sets out a NS concept that should be applied.

The EPA's EM&V guidance for EE goes into more detail about how baseline assumptions should be developed. The guidance includes definitions for GS and NS as well as baseline, applicable EM&V methods, options for using industry-standard protocols and guidelines, and other topics necessary for successfully quantifying and verifying savings. Definitions from the EM&V Guidance include:

---

[13] EPA 2015a. Environmental Protection Agency, "The Clean Power Plan," October 22, 2015.

[14] Environmental Protection Agency, "Draft Evaluation, Measurement and Verification Guidance for Demand-Side Energy Efficiency," August 3, 2013.

- **Gross savings (GS)**: savings calculated with respect to a defined baseline.[15]
- **Net savings (NS):** the difference between energy consumption with the program in place and that which would have occurred absent the program.
- **Baseline:** conditions (including electricity consumption) that would have existed without implementation of the subject EE activity. Baseline is used to estimate measure-, project- and program-related savings.
- **Common practice baseline (CPB):** default technology or condition that would have been in place at the time of project implementation absent the EE installation (EPA 2015b, p. 11).[16]

In this draft guidance, EPA notes that the definition of savings is consistent with the interpretation of GS. It also defines the term energy efficiency savings as "the difference between observed electricity usage and an appropriate 'common practice baseline' (CPB)" (EPA 2015b, p. 11), or the default condition absent the improvements, and therefore does not consider impacts outside program influence. However, absent clear state standards, NS may be used instead of savings calculated relative to a market average baseline. In such cases, the state must explain how NS are calculated and must demonstrate that savings are at least as stringent as savings relative to a market average baseline. The guidance document refers to the Uniform Methods Project Estimating Net Savings: Common Practices[17] for discussion and guidance on NS and GS.

At this stage of the review and comment process of the EM&V guidance, the EPA EM&V Guidance can be said to be consistent with what some states use to estimate GS. That said, there is a need for clarity about what EPA is proposing in the guidance.

The EPA's definition of NS follows the standard concept, and the operational definition may be the difference between observed energy use and a CPB. Some questions that might be considered include:

1) Whether the difference between the observed energy use and a CPB is in line with the baseline definition contained in the final rule.

2) Whether this approach is generally appropriate for most EE actions that would be part of a CPP (i.e., does it really get at the correct net savings concept for this purpose?).

3) Is the use of the CPB the preferred method to be used to estimate GS (and in some cases, NS) across most programs? The use of a CPB is highlighted throughout the document to a degree that might support the interpretation that, while other methods can be used, CPB is the preferred method. On page 34 of the guidance, it states: *"As indicated in Section 2.2, CPBs are the regulatory requirement for quantifying savings."*

---

[15] This a less restrictive definition of GS than is found in other EM&V guidelines where more is said about the baseline to be used in estimating GS. This definition only that a baseline is used to produce GS.

[16] See the draft Evaluation, Measurement, and Verification (EM&V) Guidance Section 2.2.2 for common practice baseline definitions for different types of energy efficiency activity (EPA 2015b).

[17] National Renewable Energy Laboratory (NRLEL), "The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures, Chapter 23: Estimating Net Savings: Common Practices," September 2014.

These questions highlight the importance of ensuring transparency (Principle #5) in methods and assumptions applied in impact estimation. The CPB approach has been used in the Northwest by the Northwest Power and Conservation Council (Council) and the regional technical forum (RTF) created to support the Council's 5 year power plans. It has not had the same level of adoption in other regions, and like all methods, it has pros and cons. It might be easier to apply in some instances and more difficult to apply in other cases. For example, developing a CPB for a commercial lighting program with a wide variety of lighting applications, lamps, and fixtures could be a complex undertaking. In addition, there are some concerns about potential bias in these estimates as people that are influenced by an EE program to take action may not be the "common" market participant. These are all discussion points that would benefit from comments.

EPA's Draft EM&V Guidance creates some challenges for policymakers: policymakers need to make decisions that they believe are meaningful for their state based on the best information available and they need to ensure that these decisions are well documented per Principle 5, *Ensure Transparency*. The EPA EM&V Guidance is broad enough in scope to encompass most approaches to produce savings estimates from EE programs. Some additional clarity is expected when the EPA responds to comments filed on the Draft Guidance.

4. Evolving Methods

GS and NS estimates can be determined using different approaches. In deciding which approach or combination of approaches to use in a given situation, several factors need to be considered. Selecting the appropriate NS analysis methods depends, in part, on the type of questions that need to be answered by a NS study and upon a number of other variables that play a critical role in determining the suitability or feasibility of each approach. These criteria include the following (Johnson Consulting 2014):

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

Another element that is increasingly being considered when making decisions about allocating evaluation efforts and resources might include the value of the information produced by the evaluation. This may be related to rigor and to the cost of the research, but value of information is in many ways its own criteria. For example, if a large program has not been evaluated in a number of years, the value of the information could be quite high independent of the other criteria. The evaluation might result in more substantive changes in program design to improve the program, and confirm the extent to which the program is providing the expected savings.

This chapter explores new and evolving methods meant to tackle some of these complexities.

4.1 Market Effects Models

The estimation of SO and ME is increasing across a number of jurisdictions and for various reasons. As resource acquisition programs mature in many markets and the acquisition cost increases, program designers are seeking alternative programming models and market transformation programming targeted at mid-stream intervention is gaining attention. As was discussed in Section 2.1, interest in moving towards market–based models is growing especially for states that are looking at redefining the distribution system business model.[18] This means that there will be an increasing need to develop metrics of success for these programs and enhance the evaluation methods to help guide these resources into appropriate EE efforts and assess the benefits and the costs of the programs.

Some experts suggest that ME can be "best viewed as spillover savings that reflect significant program-induced changes in the structure or functioning of energy efficiency markets." Prahl et al. (2013) also suggests that market transformation is a subset of ME (i.e., it is the substantive and long-lasting effects). This view implies that ME could be viewed as a subset of spillover. Regardless, they are effects that fall outside the normal calculus of tracking system savings, and new approaches are needed (DOE UMP 2014).

[18] See Section 2.3 and the New York REV discussion. There is an increased emphasis on market-based EE programs as part of this overall initiative.

This section describes three recent ME efforts which illustrate the ME work that is becoming more common. A challenge with evaluation of the net impacts of market transformation programs is identifying program attribution since information about the participants is not known. To explore methods of evaluation that align well with the programs, some more recent market transformation programs underway are building evaluation into the program design. The momentum savings method being develop by BPA and cited above should also be considered among these developing approaches.

4.1.1 EPA Retail Products Platform – Multi-Regional Evaluation[19]

The Energy Star Retail Products Platform (ESRPP) is a nationally coordinated framework for motivating retailers to change their business practices and eventually get manufacturers to produce more efficient products, with the ultimate goal to reduce the growth of electricity use attributable to plug load through lasting market transformation. By leveraging resources and similar objectives, ESRPP will make it possible for energy efficiency programs to cost-effectively motivate retailers to make changes in their stocking, promotion, and pricing practices, and to demand increasingly efficient products from their suppliers. This platform is a utility and program sponsor driven initiative with assistance from EPA for development and expansion.

What is unique about this platform is that it is a national collaborative effort to achieve scale through consistent program design—including product categories, specifications, data requirements, and midstream delivery. Coordinated programming removes duplication of effort and redundancy across neighboring service territories, thus streamlining operations, reducing costs, and allowing retailers and sponsors to focus limited resources on effective local differentiation. A more coordinated approach attempts to ensure compliance with mandated EE goals by effecting a larger market transformation. This new approach aims to be more nimble in reacting to shifts in the marketplace and much less costly than traditional programs.

While CA utilities and Northwest Energy Efficiency Alliance (NEEA) piloted this approach in 2014, a larger coordinated pilot effort led by EPA is expected to launch in 2016, [20] with opportunities for additional program participants to join in on a rolling basis. According to the EPA, the keys to the success of the ESRPP are as follows (EPA 2014c):

- Program sponsors agree to adopt a common set of retail based products for promotion.
- A sufficient number of program sponsors use the Retail Products Platform to achieve critical scale, serving more than 25% of the US population.
- Critical scale is achieved through consistent program design—including product categories, specifications, data requirements, general approach (i.e. midstream incentives), and lower per-unit incentive and administrative costs for both program sponsors and retailers.
- Retailers agree to provide unprecedented access to critical sales and market share data to program sponsors in exchange for targeted product categories and consistent and streamlined data and reporting requirements from program sponsors.

[19] This summary of market effects methods was contributed by Rick Ridge, October 2015.

[20] A complete list of states included in the 2016 and 2017 launch can be found at the following location: https://www.energystar.gov/sites/default/files/asset/document/ESRPP_1pager_10-07-15.pdf

- Retailers and program sponsors work together to tailor local go-to-market strategies built on the national framework allowing for some flexibility in local markets.
- EPA, retailers, and leading regulatory experts support program sponsors in developing and promoting supportive policy and innovative EM&V approaches.
- Retailers in the ESRPP agree to provide one year of historical and current sales data which will facilitate direct analysis and avoid the time and expense associated with gathering additional data from non-participating retailers

The simplest approach to assessing program attribution based on participating retailer data involves using a 12-month historical data series for participating retailers to forecast a counterfactual program qualified sales volume. The forecasted baseline is then compared to the actual program-period sales data post implementation. The difference between the program-period data and the forecasted baseline is the net effect of the program. The forecast will depend on a number of factors and will vary by product. In some cases, the net effects of the program may not appear until the second year or subsequent years. Evaluations of the sponsors' programs will also include tracking various short, medium and long-term performance indicators.

4.1.2 California Retail Plug-load Portfolio Program[21]

The Retail Plug-Load Portfolio Program is the first true market transformation program in California relying on a mid-stream design. In an innovative attempt to address the growing issue of residential plug-load, Pacific Gas and Electric Company (PG&E) and the Sacramento Municipal Utility District (SMUD) launched the Retail Plug-Load Portfolio (RPP) Trial in 2013 and 2014. The program design influences retailers to demand, stock, and sell a greater proportion of energy efficient models of home appliances and consumer electronics in targeted product categories to achieve greater residential customer energy and demand savings. Retailers are paid per-unit incentives for every program-qualified model that they sell during the program period. Program-qualified models are typically models that meet or exceed the minimum ENERGY STAR specification in each product category.

While the RPP Program is designed to run for ten years, the RPP Trial was conducted with limited objectives: a single participating retailer, in 26 participating stores, over the 14-month period of November 2013 through December 2014.

The RPP Program is designed to address several issues. First, while decades of EE programs have achieved tremendous energy savings, energy savings opportunities associated with plug loads have not been aggressively targeted given their relatively low per-unit energy savings. Typical downstream incentives may not be effective for electronics and appliances that have relatively small incremental savings; a small rebate to the customer will not influence the customer's decision on high end products. As administrative costs become burdensome, acceptable total resource costs for individual product measures become elusive. However, for the retailer, the small incentive may be equal to or greater than the profit margin on the purchase. This can significantly influence the retailer's promotional efforts. Economic benefits resulting from RPP Program must accrue to retailers as well as to PG&E customers in order for the program to work.

[21] This summary of the California Retail Plug-load Portfolio Program was contributed by Rick Ridge, Ridge and Associates, October 2015.

Although the energy savings opportunities associated with plug loads have relatively low per-unit energy savings, the savings potential within the plug load category is large and growing. Navigant Consulting, Inc. noted:

> *"There are many new Appliance Plug measures that are coming into the market and modeled in this study. The results show that these new appliance plug measures [computers, power strips, vending machine controllers, TVs, clothes washers, dishwashers] have a significant impact on energy savings potential and make up nearly a quarter of the [market] potential savings in 2020." (CPUC 2013, p. 153)*

Furthermore, the U.S. Department of Energy's Energy Information Administration (DOE/EIA) estimated that "other" electricity consumption combined with televisions and office equipment represented about 29% of U.S. residential annual electricity consumption in 2006 and will grow to approximately 36% in 2020. EIA forecasts that the "plug load" annual consumption will increase by 31% over the same period on a per household basis (DOE Study 2008).

As the first program relying on a mid-stream design that is a true market transformation program in California, questions remain about how to evaluate the program over the short-, mid-, and long-terms, as well as how to best implement and administer it in a cost-effective manner. The evaluation of the Trial program focused on assessing the extent to which program performance, operational, and evaluation objectives were achieved.

The program performance objectives included:

- Achieve an increase in the program-qualified share in targeted product categories among participating retailers
- Achieve gross and net energy and demand savings associated with the sales of program-qualified models in targeted product categories among participating retailers.

The seven operation/implementation objectives included, among others:

- Participating retailer develops strategic plan to sell targeted products
- Participating retailer faithfully implements the strategic plan designed to sell targeted products
- Increase retailer awareness of promoting, selling, and/or stocking products based on energy efficiency criteria.

Finally, the six evaluation team objectives included, among others:

- Test various methods to evaluate the trial.
- Analyze short- and mid-term performance indicators. Identify, operationalize, track, and analyze key proximate (short and mid-term) indicators.
- Propose performance indicators for measuring longer-term effects and/or benefits.

The evaluation team used an array of different approaches to evaluate the RPP Trial. The process evaluation surveyed retailers (merchant, marketers, and store managers), members of the retailer leadership team, and members of the implementation team. It also included a review of the salesforce database, marketing activities, retailer shelf surveys, and a review of program documents. The impact evaluation methods involved a participant-only comparison of a forecasted baseline to recorded sales, a regression analysis using program-tracking data, a simple difference-in-differences analysis involving participating and nonparticipating stores, and a review of relevant qualitative data.

Based on a preponderance of evidence approach, the evaluation of the RPP pilot found that the performance, operational, and evaluation objectives were achieved, and offered recommendations for improvements in program design and delivery, as well as future evaluation activities. The evaluation found that the short-term lift in sales for participating retailers was small. Because these are first-year savings from a small-scale trial of a new market transformation program, the savings were never expected to be substantial. Factors influencing these results are: (1) due to the short duration of the RPP Trial, only a subset of the RPP Trial activities, outputs, and outcomes could be assessed, (2) the RPP Trial promoted products that met the minimum ENERGY STAR specification, and (3) there were very limited opportunities for the retailer to make notable changes to their product assortments.

Several findings from the evaluation of the RPP Pilot are worth mentioning here as other policy makers contemplate these types of programs. First, there is a need to recognize that many aspects of the retail market will need to be transformed for the program to reach its full potential. Second, there is a notable risk to this program if regulators, utilities, and evaluators do not adequately grasp the implications tied to the program being a market transformation program. The evaluation also notes that some of the future challenges for the program and evaluation activities include that data management will continue to be a challenge as the amount of data needed to support this program is large, and that obtaining reliable data representing the nonparticipating retailers and the overall regional and national markets are big concerns.

The next step is the launch of the RPP Pilot Program in January 2016, which will add three additional retailers and expand the number of product categories. If successful, the RPP Pilot will transition to a full scale RPP Program, designed to run through 2025. Going forward, a very different approach, a theory-driven evaluation, will be used to assess its efficacy. Over time, multiple metrics will be monitored and assessed to determine whether the program is on track to achieve its ultimate objectives. Periodic reports will not contain a single estimate of a NTG ratio but rather a plausible range of NTG ratios along with an internally consistent story based on the preponderance of the evidence about the performance of the program.

4.1.3 Other Market Effects Methods – Massachusetts

In the past few years, the Massachusetts Program Administrators (PAs) have pursued a limited number of market effects studies—including the 2010 market effects study of C&I High Bay Lighting (KEMA 2014), the ME baseline characterization study for LEDs (DNV GL 2015), the Residential New Construction Net Impacts Study (NMR 2014a), and, a few years ago, the Statistical Analyses of Penetration of ENERGY STAR-compliant Appliances (Wilson-Wright 2005). Historically, market effects savings from Massachusetts' programs were infrequently counted in program evaluation for a number of reasons, including the focus on near-term savings, the complexity of assessing market effects when taking a participant-focused approach, and an incomplete understanding among stakeholders of what market effects are and how they can be measured.

In 2014, the cross-cutting evaluation team was tasked by the MA PAs to design an evaluation plan to develop methods for measuring market effects of MA's EE Programs. The objectives of this study were to:

- Describe the key concepts important to understanding and measuring market effects and associated savings
- Help the PAs identify when they should consider measuring market effects by describing the conditions likely to produce substantial market effects
- Identify conditions allowing measurement of savings from market effects
- Delineate and describe the range of methods available to measure market effects and provide guidance for selecting among them when planning for evaluation.

As part of this study, the evaluation team delivered two workshops with the PAs to achieve a mutual understanding of the key concepts and ME and how to measure market effects and associated savings (Principle #1: *Establish a Common Understanding*). A final report delivered in November 2014 (NMR 2014b) provided the following key findings of when programs are more likely to result in substantial market effects:

- When the savings per transaction are small, but the transactions are numerous.
- When the program strategies in use are likely to result in market changes (e.g., the programs target markets rather than program participants)
- A significant proportion of market actors have been touched by the program
- The product or service that the program addresses offers significant non-energy benefits.

Other key findings included in the final report include:

- Theory-based evaluation is an important planning tool for ME measurement and a qualitative method for determining if ME have resulted from program efforts. Theory-based evaluation techniques are useful both for assessing savings from past ME from established programs (retrospective evaluation), and for laying the groundwork for future assessment and documentation of ME (prospective evaluation).
- Theory-based evaluation should serve as the framework on which all ME evaluations are based regardless of the specific analytical method this is used to establish quantitative attribution of market effects for a particular program.
- Theory-based evaluation requires advance planning. Key activities to undertake when planning include:
  - Identify the markets that the program targets
  - Characterize the market to get an accurate understanding of the market as well as the market actors
  - Develop a market model
  - Tell a story (the program theory)
  - Develop a logic model

- Establish indicators tied to expected ME outcomes (along with the indicators, identify their data source and timing of data collection)
- Identify baselines.

- There are four general methods of estimating net savings stemming from ME:
  - Supply-side market actor self-reported counterfactual analysis (self-report surveys/interviews with upstream market actors about free ridership and spillover)
  - Cross-sectional analysis by identifying one or more comparison groups where sales will be tracked along with the program area
  - Forecasting or retro casting the non-intervention baseline which involves developing a statistical model to estimate how the market would behave over time without the intervention of the program
  - Structured expert judgement which involves identifying a team of experts who review information on the market (e.g., Delphi panels).

This study is being used to inform and plan future ME evaluation efforts in several markets (e.g., HVAC, lighting) in MA.

4.2 Bottom-up versus Top-down applications at the state, regional and national levels[22]

Regulatory agencies and investor-owned utilities (IOUs) have begun to explore “top-down” analysis as a supplemental or alternative approach to measuring net energy program impacts. The analysis looks at energy consumption over time across geographic areas to estimate the effect of program activity on consumption while controlling for other factors that vary over time and areas. This analysis is an econometric model using aggregate (e.g. state, portfolio, or sector-level) cross-sectional and time series consumption and econometric data.

The method is referred to as “top-down” because it extracts the overall program portfolio effect from a decomposition of total aggregate consumption. In contrast, traditional “bottom-up” approaches typically estimate savings at the individual measure or program level, then sum these to develop total portfolio impacts.

Bottom-up approaches use a variety of methods, including deemed savings based on engineering computations, on-site observation and monitoring, engineering analysis of varying levels of complexity, and premise-level consumption data analysis. Net savings adjustments are designed to account for free-riders, spillover, and snapback. These adjustments are sometimes embedded in consumption-data analysis, but more commonly involve survey data collection in some form. Concerns related to bottom-up net savings estimates include the data collection costs, the difficulty of obtaining robust estimates of some effects, and the dependence of net-to-gross adjustments on methods and assumptions.

[22] This summary of top-down methods was provided by Dr. Miriam Goldberg and Noel Stevens of DNV GL. DNV GL is leading the exploration of top-down methods in MA as part of the cross-cutting evaluation team.

Top-down approaches are of interest for their potential to mitigate the challenges of bottom-up methods; however, they face similar challenges. Top-down models measure changes to aggregate energy consumption relative to changes in energy efficiency programmatic activity, prices, and other economic factors. The goal of this type of modeling is to isolate the effect of program activity from other natural changes and policy variables. Top-down techniques use a holistic approach by estimating 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). The appeal of this structure is that, in principle, it captures the full program effect, including FR, SO, ME, and snapback. A properly structured top-down model can potentially make the following contributions to the set of tools used to evaluate energy efficiency programs:

- Provide relatively inexpensive estimates of program-induced savings estimates for all geographic units in the study comparted to the combined bottom-up approaches across the region
- Provide expected program-induced savings for a geographic unit with particular characteristics
- Provide combined effects of all cumulative program activity for a particular geographic unit, including spillover and snapback.
- Provide confidence intervals and precision levels for net energy savings from the portfolio of programs.

Despite the potential advantages, the following challenges constrain the ability to develop robust top-down models:

- Models face substantial data limitations resulting in compromise between the ideal specification and the types of data available at various levels of aggregation
- Spillover between study units reduces apparent program effects, limiting the model's ability to isolate true program impacts.
- It is nearly impossible to account for all factors that influence consumption, particularly given the data limitations, so that model results are potentially biased by omitted or incorrectly specified variables, or incorrectly specified model forms
- In some circumstances, self-selection of participants into programs can bias the estimated net effects, as with bottom-up methods
- Alternative reasonable model specifications may lead to substantial differences in estimated effects, particularly if shorter time series are used
- Fitting a model across a longer time series requires consistency over an extended time in the overall pattern of how the non-program and program variables affect consumption.

These technical constraints have prevented top-down models from obtaining true estimates of savings net of free riders. The approach is also limited in the type of information it can provide, even if highly accurate. Top-down methods cannot provide separate FR, SO, and ME estimates by program or measure type. The models provide average program effects across the time frame and areas of the analysis, not individual program-year effects. Therefore, these models cannot be used to isolate the relative contribution of individual programs or measures to overall savings, or indicate year-to-year changes in program effectiveness. The methods cannot identify which groups of measures are performing better, or worse compared with others with similar characteristics. Such information is valuable to policy makers, program designers, and implementers.

Even if top-down models cannot provide a robust, precise measure of portfolio net savings, they nonetheless can potentially offer a high-level empirical demonstration that program effects are real and ongoing. This perspective can be valuable for system operators and planners, as well as program staff and regulators. Thus, top-down methods may provide an additional (supplemental) tool in the set of tools used to evaluate the portfolio of programs, provided the data limitations can be addressed.

Utilities and regulatory agencies can work toward developing a platform for estimating effective top-down models by maintaining historical consumption and program tacking data at the individual account level across many years. These data are typically not retained for more than 3-5 years, and do not capture data sufficient to properly account for the cumulative effects of programs over time, nor to capture the necessary diversity in programmatic activity.

4.3 Common Practice Baselines

Another way to examine baselines for GS and NS involves the development of “common (or current) practice baselines (CPB)”, “industry standard practice (ISP),” and “business as usual (BAU)” assessments of the market. These approaches are not new, and have been used in estimating energy savings back to the 1960s. However, they have been receiving increasing attention in recent years. For example, the Draft EPA EM&V Guidance for EE Savings (EPA 2015b) discusses CPB approaches for the development of Technical Resource Manuals (TRMs), and sets out procedures for initial estimates of energy savings using these approaches. This section is not meant to advocate a specific way to set a baseline, but to present a few issues that that may influence how these concepts fit into an EE program or portfolio evaluation framework. There are several ways to use CPBs in GS and NS estimation and there are policy views that can support these different options. Three options are presented below:

1. **Use a CPB as the basis for calculating energy savings and use those savings estimates as the final values for program savings with no adjustment.** There are some arguments in favor of doing this. In the end, it depends on what assumptions you believe and what your jurisdiction wants the savings to represent from a policy perspective.[23]

2. **Use a CPB, but make adjustments for potential biases (e.g., self-selection) or omitted savings.** “Common practice baselines are estimates of what a typical consumer would have done at the time of the project implementation. Essentially, what is “commonly done” becomes the basis for baseline energy consumption” (SEE Action 2012, p. 7-2).[24] However, the consumers (or participants) that take action as a result of a program may not be well represented by the “typical consumer.” If the program is an opt-in choice for participants,

[23] Two sources that set out used of the common practice baseline are: 1) Eckman, T. “Savings Are What You Define Them to Be: The ‘Baseline’ Issue,” presentation at the RTF Meeting, December 17, 2013. 2) Regional Technical Forum, “Roadmap for the Assessment of Energy Efficiency Measures” (RTF 2014).

[24] SEE Action illustrates this “commonly done” baseline using an appliance example. “For example, if the program involves incenting consumers to buy high-efficiency refrigerators that use 20% less energy than the minimum requirements for ENERGY STAR® refrigerators, the common practice baseline would be refrigerators that consumers typically buy. This might be non-ENERGY STAR refrigerators, or ENERGY STAR refrigerators, or, on average, something in between” (2012, p. 7-2).

self-selection bias may be a concern. Also, there does not appear to be a mechanism by which to count SO or ME as part of the calculated by the CPB. Again, this is a policy question and it involves the symmetry principle regarding the consideration of NS factors that might lower savings estimates and those that might increase savings estimates.

3. **Use CPBs as initial estimates of GS and adjust them as appropriate for net savings factors (FR, SO and ME) to get at an estimate of NS.** This approach basically takes the position that there are no shortcuts and that a symmetrical analysis of the positive and negative potential biases is needed. If one uses a control/comparison group approach with matching and tests for whether the assumption of randomness is reasonable, then there may be no need to separately estimate GS as the statistical analyses of the participant and comparison groups directly produce NS estimates. In some cases, they may capture short-term participant spillover, but they have the same weaknesses as the other methods in that they don't capture non-participant spillover and market effects. These have to be estimated using additional analyses; however, it is important that double counting savings adjustments in GS and NS estimation be avoided. This re-enforces the need for clear definitions at the conceptual and operational levels.

4.3.1 Context on CPB Options

There are views that can be consistent with each of the options above. One part of the decision includes what a jurisdiction wants the savings estimates to represent.

A view that might be consistent with Option 1 above might be based on the argument that previous EE programs have affected current markets. Due to past programs, more EE equipment may be standard in the market today through spillover and market effects that have occurred over the years, but not counted in evaluations.[25] This results in current common practice baselines that are more efficient than they would have been. In this case, the CPB or market average may contain some participants (for example, end users, installers, and distributors) and nonparticipants who have been influenced by past program years. As a result, a jurisdiction may view savings accruing today as partially due to programs in previous years, but not counted. These savings along with the savings from current programs serve as a reasonable estimate of EE program impacts over the long term. From this perspective, an argument can be made that the common practice approach best represents the overall return on investment in EE.

Alternatively, a jurisdiction may focus on EE as a resource, i.e., a decision to invest additional ratepayer funds to help meet future needs for energy services. When looking at EE programs as producing new kWh resources (i.e., megawatts), a jurisdiction may take the position that each EE program should be evaluated as an incremental investment. A program implemented in 2016 should be evaluated against what is attributable to that investment only -- impacts from prior years' programs are essentially sunk costs and should not be considered.

The need for adding market effects to CPBs is being considered by the Bonneville Power Administration (BPA):

[25] An important part of this argument is that these savings have not yet been counted, and should be reflected in current savings estimates for assessing the overall contribution of EE.

> *"Energy efficiency occurs outside of direct program incentives all the time. Some of it we call spillover; some of it we call market effects; some of it we call codes and standards.* ***Momentum Savings are all the energy efficiency that occurs outside of programs, regardless of why and how****. … BPA tracks these savings because they have become a significant energy efficiency power resource in the region. Momentum Savings help BPA and the Northwest meet energy efficiency targets, define program baselines and understand what efficiency is still available." (BPA 2015). http://www.bpa.gov/EE/Utility/research-archive/Pages/Momentum%20Savings.aspx*.

BPA has already completed several of these Momentum Savings studies. This is an example of where policy and analytic views of NS estimation are linked.

There have been questions about the cost and complexity of CPB methods versus traditional methods. All methods pose challenges, and costs will vary across program type and sector. For example, it can be difficult to gather data on common practice for a large number of measures across different types of customers. There is also a need to updating CPBs as needed. Attempts to gather actual EE measure sales have proved difficult, and extrapolation from a small base of sales data to the jurisdiction or service territory can be uncertain. Trade ally panels can be used to develop expert-based judgmental baselines. However, the scale of the effort is important to assess. There may be a large number of measures and practices that need to have CPBs developed and updated with reasonable frequency. It is not clear that a general statement can be made about whether CPB methods are any more or less resource intensive than other methods, considering the level of accuracy desired in the savings estimates.

5. Other Issues in Evaluation

There are several other trends that are worth noting. 1) Increasing Availability of Data (i.e., Big Data), and 2) the Expanding Role of Random Control Trials.

5.1 High Frequency Data and More Data

The first trend is the rapid expansion of data from smart meters for evaluation and the issues that it raises. There is no question that more data is available at a faster rate than ever before. How to use these data productively poses challenges. There is also the promise by some of "real-time" evaluation, but the right question is what time period is appropriate for evaluation. If a new AC unit is installed at a customer's home at noon in May, is it important to know what the energy use was at 1 pm that same day? Probably not. Instead, information on the load reduction on peak days might be important, or energy savings on a monthly and seasonal basis. Knowing consumption hour-by-hour can help with load shapes, but to estimate these reliably, a reasonable amount of data needs to be gathered over an appropriate amount of time – this type of information can produce results in a more timely fashion in the future, but likely not in "real-time." As a result, the promise of "real-time" evaluation needs to be considered carefully.

There are real benefits that can be produced. A number of the benefits likely involve program design and implementation rather than the evaluation function. Audits can be conducted on homes and commercial buildings without site visits, and this information can be used to help target appropriate programs for customers, even if the audit results are a bit rough due to the lack of on-site information. In addition, the flow of data might help augment tracking system estimates to provide up-to-date information, but it seems unlikely that this will displace the role of evaluation that occurs after a period of time to verify program savings using appropriate in-field processes. One question is when do you really need the information? The tracking system gives you weekly data on installations and the best available information on energy savings, and that happens now. High-frequency data may improve this process, but it is not clear how much the improvement might be. Market research and process evaluations are still important for planning. This is just a snapshot of a few issues. [26,27]

5.2 Expanding Role of Random Control Trials

Most of the evaluations conducted have been limited in terms of addressing certain types of bias. Self-selection bias has been of most concern. This can be solved by randomly assigning customers to a participant group and a control group. In the past, regulators have been wary about not allowing all customers access to an EE program due to equity considerations. This comes at a price in terms of being able to accurately evaluate a program. Recently, there has been more openness to consider random control trials. Several utilities have used random control trials to test out new time-differentiated rates, and it is becoming common to use random control trials in assessing home energy usage report programs. In addition, new matching methods are allowing better quasi-experimental designs where we can get several "like" matches for each participant in a program. This is also a recent trend in EE impact studies and improvement in methods. These matching processes can benefit from having hourly data on the entire population of customers.

There have been calls for regulators to more actively cooperate with evaluators in the design of evaluations. Benefits can come from this if more information can be obtained from studies where randomization is part of the evaluation design.[28]

---

[26] A more complete discussion of this topic can be found in Violette, D., "The Promise of Evaluation 2.0," AESP Magazine, Premier Issue, 2015, available at:http://www.navigant.com/~/media/WWW/Site/Insights/Energy/2015/AESPMagazine_EvaluationDViolette.ashx

[27] LNBL has launched a project on M&V 2.0. Information on this can be found at http://energy.gov/eere/buildings/downloads/assessment-advanced-measurement-and-verification-methods-mv-20. In addition, LBNL has been collecting information on trends in evaluation – see: http://eetd.lbl.gov/news/events/2014/05/29/emv-20-what-s-the-future-of .

[28] See the section on random control trials in DOE UMP (2014).

6. References

ACEEE 2015. American Council for an Energy-Efficient Economy, “Market Transformation,” 2015, available at: http://aceee.org/portal/market-transformation.

CPUC 2013. Navigant Consulting, Inc., “2013 California Energy Efficiency Potential and Goals Study,” prepared for the California Public Utilities Commission, November 26, 2013, available at: http://www.cpuc.ca.gov/NR/rdonlyres/29ADACC9-0F6D-43B3-B7AA-C25D0E1F8A3C/0/2013CaliforniaEnergyEfficiencyPotentialandGoalsStudyNovember262013.pdf.

DNV GL 2015. MA LED Market Effects: Baseline Characterization, prepared for the Massachusetts Program Administrators and the EEAC, March 1, 2015.

DOE UMP 2014. National Renewable Energy Laboratory (NRLEL), “The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures, Chapter 23: Estimating Net Savings: Common Practices,” September 2014, available at: http://www.energy.gov/sites/prod/files/2015/02/f19/UMPChapter23-estimating-net-savings_0.pdf.

EPA 2014a. Environmental Protection Agency, “Carbon Pollution Emission Guidelines for Existing Stationary Sources: Electric Utility Generating Units,” June 18, 2014, available at: https://www.federalregister.gov/articles/2014/06/18/2014-13726/carbon-pollution-emission-guidelines-for-existing-stationary-sources-electric-utility-generating.

EPA 2014c. Environmental Protection Agency, “The Energy Star Retail Products Platform,” October 21, 2014, available at: https://www.energystar.gov/ia/partners/downloads/meetings/2014/ESRPP_1pager_Final.pdf.

EPA 2015a. Environmental Protection Agency, “The Clean Power Plan,” October 22, 2015, available at: http://www2.epa.gov/cleanpowerplan.

EPA 2015b. Environmental Protection Agency, “Draft Evaluation, Measurement and Verification Guidance for Demand-Side Energy Efficiency,” August 3, 2015, available at: http://www2.epa.gov/cleanpowerplantoolbox/draft-evaluation-measurement-and-verification-guidance-demand-side-energy.

EPA 2015c. Environmental Protection Agency, “FACT SHEET: Energy Efficiency in the Clean Power Plan,” August 20, 2015, available at: http://www2.epa.gov/cleanpowerplan/fact-sheet-energy-efficiency-clean-power-plan.

KEMA 2014. KEMA, “Final Report HBL Market Effects Study Project 1A New Construction Market Characterization,” prepared for the Massachusetts Energy Efficiency Program Administrators, June 7, 2011, available at: http://ma-eeac.org/wordpress/wp-content/uploads/High-Bay-LIghting-Market-Effects-Study-Final-Report.pdf.

NEEP 2010. NMR Group, Inc. and Research Into Action, Inc., “Net Savings Scoping Paper,” prepared for Northeast Energy Efficiency Partnerships: Evaluation, Measurement, and Verification Forum, November 13, 2010, available at: http://www.neep.org/net-savings-scoping-paper-1.

NEEP 2012. NMR Group, Inc. and Research Into Action, Inc., “Regional Net Savings Research, Phase 2: Definitions and Treatment of Net and Gross Savings in Energy and Environmental Policy,” prepared for

Northeast Energy Efficiency Partnerships: Evaluation, Measurement, and Verification Forum, December 4, 2012, available at: http://www.neep.org/net-savings-research-phase-2-0.

NEEP 2013. Northeast Energy Efficiency Partnerships: Evaluation, Measurement, and Verification Forum, "Model EM&V Methods Standardized Reporting Forms," 2013, available at http://www.neep.org/initiatives/emv-forum/model-emv-methods-standardized-reporting-forms.

NEEP 2014. Synapse Energy Economics, Inc. "Cost-Effectiveness Screening Principles and Guidelines," prepared for Northeast Energy Efficiency Partnerships: Evaluation, Measurement, and Verification Forum, November 2014, available at: http://www.neep.org/file/2873/download?token=f4VbVWAH.

NEEP, 2014. Johnson Consulting, "Decision Framework for Determining Net Savings Approaches," prepared for Northeast Energy Efficiency Partnerships Regional Evaluation, Measurement and Verification Forum, September 2014.

NEEP 2016. Gross Savings and Net Savings: Principles and Guidance, April 2016.

NMR 2014a. NMR Group, Inc., "Massachusetts New Construction Net Impacts Report," prepared for the Massachusetts Electric and Gas Program Administrators, January 27, 2014, available at: *http://ma-eeac.org/wordpress/wp-content/uploads/Residential-New-Construction-Net-Impacts-Report-1-27-14.pdf*.

NMR 2014b. NMR Group, Inc., "Methods for Measuring Market Effects of Massachusetts Energy Efficiency Programs," prepared for the Massachusetts Electric and Gas Program Administrators, November 2014.
NY PSC 2015a. State of New York Public Service Commission, "Order Adopting Regulatory Policy Framework and Implementation Plan," February 26, 2015, available at: http://documents.dps.ny.gov/public/Common/ViewDoc.aspx?DocRefId={0B599D87-445B-4197-9815-24C27623A6A0}.

NY PSC 2015b. State of New York Public Service Commission, "Staff Proposal: Distributed System Implementation Plan Guidance," October 15, 2015, available at: http://documents.dps.ny.gov/public/Common/ViewDoc.aspx?DocRefId={F3793BB0-0F01-4144-BA94-01D5CFAC6B63}.

RTF 2014. Regional Technical Forum, "Roadmap for the Assessment of Energy Efficiency Measures," June 17, 2014, available at: http://rtf.nwcouncil.org//subcommittees/Guidelines/RTF%20Guidelines%20(revised%206-17-2014).pdf.

SEE Action 2012b. Schiller Consulting, Inc., "Energy Efficiency Program Impact Evaluation Guide," prepared for State & Local Energy Efficiency Action Network, December 2012, available at: http://www4.eere.energy.gov/seeaction/system/files/documents/emv_ee_program_impact_guide_0.pdf.

TRM MP 2015. Joint Utilities, "Technical Resource Manual Management Plan," June 1, 2015, available at: http://documents.dps.ny.gov/public/Common/ViewDoc.aspx?DocRefId=%7B9568AF5D-1484-4CA6-8EB7-C2334DF673E1%7D.

Violette et al., "Final Report: Iowa Energy-Efficiency Net-to-Gross Report, Prepared for: Iowa Utility Association and the Oversight Committee," by Navigant Consulting, June 2015.[1]
https://efs.iowa.gov/cs/groups/external/documents/docket/mdax/mjax/~edisp/1201492.pdf.

Wilson-Wright 2005. Wilson-Wright *et al.*, "Front-loading Marketing: Assessing Cumulative Effects of ENERGY STAR® Appliance Promotions on State-by-State Market Penetration Levels," 2005, available at: http://www.iepec.org/conf-docs/papers/2005PapersTOC/papers/079.pdf.

[1] Other links to this work including the Commission Order and comments by parties can be found at:
Commission Order:
https://efs.iowa.gov/cs/groups/external/documents/docket/mdax/mziy/~edisp/1322686.pdf
Comments by Parties:
https://efs.iowa.gov/cs/groups/external/documents/docket/mdax/ndy3/~edisp/1467318.pdf
https://efs.iowa.gov/cs/groups/external/documents/docket/mdax/ndy3/~edisp/1467085.pdf
https://efs.iowa.gov/cs/groups/external/documents/docket/mdax/ndy2/~edisp/1466889.pdf.