Net Savings Scoping Paper

FINAL

11/13/2010

Submitted to:

Northeast Energy Efficiency Partnerships:
Evaluation, Measurement, and Verification Forum

Submitted by:

NMR Group, Inc.
Research Into Action, Inc.

50-2 Howard Street, Somerville, MA 02144
Phone: (617) 284-6230 Fax: (617) 284-6239
www.nmrgroupinc.com

Abstract

The EMV Forum requested the preparation of a scoping paper with the purpose of improving Forum members' understanding of how net energy savings is defined, how stakeholders use net savings, and the opportunities and barriers to increasing the consistency of and quality in net savings definitions and measurement in the region. Three issues motivated the request for the study: 1) the prevailing lack of consistency in definitions and measurement in the Northeast; 2) expanded use of energy savings estimates by diverse audiences, particularly with applications to climate change policies; and 3) increasing challenges of determining program "attribution"—that is, demonstrating that an energy efficiency program in a given year caused savings to occur in the face of extensive prior program activity and the existence of additional influences promoting efficient actions. The paper explores these topics through a literature review of over 100 sources, interviews with 12 experts on energy efficiency programs and air regulation, and feedback from Forum members. The paper captures the perspectives of energy and air regulators, program administrators, and energy efficiency and other social science evaluators.

The paper finds that the energy efficiency community persists in using different definitions of net savings, although the most common definitions emphasize the concepts of free ridership and spillover. Other potential concepts—rebound, leakage, realization rates—are sometimes considered as part of net savings and other times as adjusted gross savings. The paper also highlights the arguments for and against the measurement of net savings. Arguments for net savings focusing on determining actual program impacts and insuring that ratepayer and taxpayer funds are spent responsibly; arguments against measuring net savings highlight challenges with measurement methods and the focus on free ridership at the expense of other program impacts. The paper discusses the key advantages and disadvantages of current net savings methods, addressing issues such as validity (especially construct validity), reliability, and bias as well as the psychological and behavioral processes that affect how people respond to questions designed to measure free ridership and spillover. The scoping paper examines the ways in which approaches to net savings do and do not meet current and evolving policy needs, with the latter focused especially on the needs of air regulators who may rely on adjusted gross instead of net savings. Finally, the paper examines the advantages and disadvantages of pursuing a consistent regional approach to net savings and the potential challenges of implementing consistency.

The paper concludes that, without a policy driver, Forum members have no motivation to increase consistency in net savings definition or methodologies. Furthermore, the air regulation and energy efficiency communities do not share a common language or understanding of what might be needed to translate energy savings into emissions reductions. The paper offers numerous recommendations related to adopting consistent definitions of adjusted gross savings, net savings, and attribution, advocating for higher-quality data for use in methods to estimate net savings, moving toward consistency in approaches to net savings, and identifying ways of translating energy savings into emissions reductions through methods that are mutually acceptable to the energy efficiency and the air regulation communities.

Contents

Tables

Figures

Executive Summary

NMR Group, Inc. and its subcontractor Research Into Action, Inc. (the project team) are pleased to submit this scoping paper on Net Savings to Northeast Energy Efficiency Partnerships, Inc. (NEEP) and its Evaluation, Measurement, and Verification (EM&V) Forum (hereafter EMV Forum or just Forum).[1] The goal of the project is to produce a scoping paper that informs and supports the EMV Forum goals of understanding the reasons for measuring net savings, increasing the consistency and quality in EM&V practices with respect to defining and estimating energy efficiency program net savings within the Northeast Region, and identifying the needs of stakeholders (*e.g.,* air regulators[2]) who might (or might not) use net savings estimates to measure progress toward non-energy goals.

The energy efficiency community pursues the estimation of net savings because the concept separates the energy savings directly caused by the ratepayer or publicly funded program from those savings that would have happened anyway. The purpose underlying net savings estimation is to establish whether the program represents a wise use of funds. Estimating net savings, however, necessitates estimating a counterfactual, or something that did not happen—in this case, what energy savings would have been in the same context but without the program. As discussed below in the executive summary and the main report, methodologically, this charge is extremely difficult—some would say impossible—to fulfill, and the results always embody a degree of uncertainty because we can never know for certain what would have happened without the program. Despite the challenges of measurement, estimation of net savings is critical to energy efficiency program assessment; without net savings estimates we can never truly be certain that a program or portfolio has an effect. The current and evolving contexts in which programs operate, however, may mean that the time is ripe for changing the common approaches to estimating net savings. This scoping paper summarizes the debates surrounding the importance of measuring net savings, the challenges inherent in such measurement, and possible directions for the near and more distant future.

The Forum was motivated to sponsor this scoping paper for three primary reasons, discussed as follows.

Exploration of Consistent Approaches to Net savings

The first motivation for this study was to explore the possibility of consistent definitions of and approaches for measuring net savings in the Northeast. Across the region, jurisdictions conceptualize "net energy savings" differently, particularly with respect to free ridership and spillover. Likewise, program administrators and evaluators rely on numerous methodologies to

[1] This executive summary is intentionally more comprehensive than the typical executive summary to provide policy makers and others with limited time to read the full report enough information to understand the recommendations

[2] For the purposes of this papers, "air regulators" refers to federal and state environmental regulators focused on greenhouse gas emissions.

estimate net savings, but these program administrators and evaluators often voice at least some concerns about reliability, validity, and bias.[3] The diverse range of definitions of net savings, and methodologies for estimating net savings, presents challenges for inter-jurisdictional efforts to meet goals for energy and greenhouse gas emission reductions, as savings being reported are neither consistently comparable nor equally valid or reliable. Moreover, differences in ways of defining and measuring net savings also increase evaluation and reporting needs for program administrators operating in multiple jurisdictions as well as for market-level programs, which are expected to have effects beyond the immediate jurisdiction in which they operate. The Forum sought to assess the possibility of developing a consistent regional approach to defining and measuring net savings across the region, and to develop an understanding of what this approach might look like if it were to be developed and adopted across the Northeast.

Expanded and Diversified Audiences for Net Savings

The second motivation for this study stems from the recent expansion and diversification of the audiences for net energy savings. More jurisdictions now have efficiency programs, the savings targets of existing programs have been recently expanded, and energy efficiency programs are expected to provide substantial reductions in greenhouse gas emissions, now being regulated as an air pollutant by the Environmental Protection Agency (EPA). For these reasons, some existing audiences for net savings estimates are paying closer attention to their definitions and ways of measuring net savings, while new audiences are beginning to ask difficult questions. The Forum was interested in understanding the extent to which current ways of defining and measuring net savings meet the needs of the diversifying audiences, and the extent to which greater consistency would improve their ability to use the estimates to meet their energy use and greenhouse gas reductions goals.

Increasing Challenges in Establishing Attribution and Causation

The third motivation for this study relates to the concept of "attribution." Attribution refers to the practice of determining what impacts are *caused* by a specific program during a specific time period—usually one or a few program years, and occasionally over a longer period, reflecting the fact that previous program activity can affect current net savings. Some people view attribution and net impact assessment as the same thing; the terms can be used interchangeably. For others, including most of the individuals we interviewed, attribution analysis involves the measurement and tracking of *proximal*—or performance-related—measures and determining if they are caused by the program rather than by other factors that may lead to the same outcome. For example, one might assess whether the number of advertisements featuring the ENERGY STAR® logo

[3] In research communities, "validity" refers to whether the estimate measures what it was intended to measure, "reliability" refers to the ability of the method to produce the same results in repeated trials, and "bias" refers to whether the estimate differs from the "real" value in a systematic manner. These three terms are often used interchangeably in common nomenclature, but we have attempted to adhere to these more exact definitions in this report to avoid any confusion regarding the terms.

increased after a retailer made use of cooperative advertising funds and if this differed from a comparable media market in which retailers did not make use of the funds. Sometimes attribution assessment takes the additional step of estimating free ridership and spillover so as to measure the savings that occurred only because of the program; this gets closer than do proximal measures to assessing the *ultimate* program goal of reducing energy consumption.

Attribution assessment has always involved isolating the effects of the program from other influences. Increasingly, however, when the energy efficiency community mentions the "challenge of attribution" or "sorting out attribution," it refers to the fact that reductions in end users' energy consumption can be affected not only by myriad efficiency programs offered by a broad range of sponsors, but also by economic ups and downs, changes in energy prices, concerns about climate change, and ongoing advances in technology, among other influences. This situation has significantly exacerbated the difficulty of establishing causation, and, therefore, of estimating net savings, and it is likely that this situation will persist. Because of the increased difficulty of establishing causation, some commentators in the energy efficiency community believe that the net savings estimates developed recently are less accurate those developed in the past when there were fewer programs and messages promoting efficiency and "being green". Therefore, the Forum wanted to explore how the increasing challenges associated with attribution may affect net savings definitions and methodologies.

Primary Project Objectives

To address these issues of consistency, audiences, and attribution, the specific project objectives are as follows:

- To identify the key audiences that rely on net savings to meet policy and program needs
- To highlight the reasons for measuring net savings and the situations in which such measurement is and is not appropriate
- To document the issues and challenges associated with existing approaches to defining and measuring net savings
- To propose recommendations for next steps needed to make progress toward the goal of increasing the consistency and quality of definitions and methods related to net savings

Methods

The project team gathered information for this scoping paper using two separate methodologies. The first involved a literature review of approximately 100 articles, papers, presentations, and book chapters from within the field of energy efficiency, as well as other evaluation fields. The literature review serves two important purposes. First, it provides a context—over time, across locations, and beyond energy efficiency—for issues related to program impacts generally and net savings in particular. Second, the process of conducting the literature review informed the in-depth interviews, by pointing to issues to address with the experts. In conducting the review, we

documented the various themes and perspectives on net energy savings represented in the readings, paying particular attention to questions related to energy and climate change policy.

We also interviewed 12 experts on the administration, implementation, and evaluation of energy efficiency programs for energy regulation and air regulation in order to gain insight into issues related to energy savings in general and net savings in particular.[4] While the literature review provided the team with an understanding of the dominant concerns, issues, and viewpoints related to net savings, the in-depth interviews allowed us to focus more specifically on the issues of greatest concern to the Forum that have not been adequately addressed in the existing literature—specifically the issues of the expanding audiences for net savings estimates and current and potential uses to which the estimates may be applied. The Forum designated a project subcommittee consisting of Forum members to assist NEEP and the project team. The subcommittee and the project team worked together to identify potential interviewees representing a mixture of energy regulators, air regulators, program administrators, and widely acknowledged experts on efficiency programs, net savings, and air regulation. We fielded two interview guides: one for individuals more familiar with energy efficiency programs and net savings, and another targeted at air regulators, who have varying levels of experience with the concept of net savings. Appendix A includes the two interview guides.

Finally, after the project team had submitted a draft report to the Forum and the subcommittee members had provided feedback on the report and also had shared their own opinions on the topics addressed in the study, NEEP directed the project team to treat such comments as data for inclusion in the study. Although many of the comments have been circulated to all subcommittee members, we still treat the material as confidential by not identifying the individuals who made the observations, as this paper will be shared beyond the Forum.

Net Savings: Contextual Issues

In conducting this research, the project team identified three key issues that set the context for the exploration of net savings measurement approaches, existing and evolving policy needs related to net savings, and consistency in net savings evaluation. We also explored how other evaluation fields address net impacts.

Meeting the Needs and Expectations of Diversifying Audiences for Net Savings

As noted above, the diversifying and expanding audiences for energy savings estimates—gross and net—was one of the motivations for this project. In fact, the project team found this issue to have a strong influence on the results of our research efforts. Historically, program administrators and energy regulators have been the main audiences for net savings estimates,

[4] The independent system operators (ISOs) in the Northeast have decided to accept estimates of adjusted gross savings for the Forward Capacity Markets. For this reason, we did not include system planners among our interviewees. We recognize that their perspectives and needs may be relevant to consider in the future—for example, in the context of incorporating energy efficiency into system planning forecasts.

with program planners as an important secondary audience. These groups have used net savings estimates to assess how well programs were performing, to guide program revisions and discontinuation, and to decide on rewards or penalties for program administrators.

More recently, the audiences for energy savings estimates have expanded to include stakeholders and regulators with expectations that energy efficiency efforts will realize substantial reductions in greenhouse gas emissions, while long-standing audiences have increased their scrutiny of the estimates, largely in response to increased funding and expanded regional goals for energy savings and reduced emissions. While the existing and emerging audiences share some of the intended uses of energy savings estimates, their differing needs and perspectives contribute to differences of opinion regarding whether they require estimates of gross or net savings, and, if so, how to define and measure net savings, whether those definitions and methods should be consistent across the Northeast Region, the importance placed on net savings in assessing program performance, and even whether net savings should be measured at all. In short, the increased diversity of audiences is an underlying consideration in each of the topics discussed in the paper. Air regulators are one of the most critical new audiences for estimates of energy savings resulting from efficiency programs, although it is unknown at this point if they will require adjusted gross savings or net savings; therefore, in this report we pay particular attention to questions and concerns raised by air regulators.

Definitions and Conceptualizations of Net Savings

How one defines net savings is the second contextual issue. Sources from within the energy efficiency community unanimously agree that net savings are those that would not have occurred without the program.[5] Yet the operational definition of net savings differs among programs and jurisdictions. The literature and interviewees largely addressed two components of net savings: free ridership and spillover. Jurisdictions differ in their approach to net savings overall and these two components in three different ways. First, some jurisdictions allow for the inclusion of both concepts in their definitions of net savings, while others allow only free ridership to be counted. Second, some jurisdictions expect free ridership and spillover to be isolated and measured, but others allow estimates of net savings in which free ridership and spillover are not individually measured but are nevertheless embedded in the estimate (as in many market-level estimates of net savings). Finally, a few jurisdictions in the Northeast do not require the measurement of net savings—or its components—at all; instead they rely on gross savings estimates.

[5] Although the concept of net savings appears to align to the concept of "additionality" in the air regulation community, this might not actually be the case in practice. Additionality is defined by the Forum as "A criterion that says that avoided emissions should be recognized only for project activities or programs that would not have 'happened anyway' in relation to a baseline estimate of project activity and associated with emissions reductions" (Horowitz 2009). The World Resources Institute (2005) states that how this concept is put into practice remains a source of debate and confusion. It is our understanding that the EPA is developing a roadmap for what the energy efficiency community will need to do to have its savings counted towards emissions reductions. To the extent that the roadmap addresses additionality, its approach may or may not align with the concept of net savings in energy efficiency.

Here it is important to note that, while nearly all interviewees and literature sources recognize that net savings is not synonymous with free ridership, many commentators perceive that energy regulators focus almost exclusively on free ridership, making it the *de facto* measure of concern. Therefore, one contextual theme underlying many of our findings is a frustration among some commentators—and an agreement among others—with this focus on free ridership.

Some sources also named other possible adjustments to gross savings (*e.g.*, realization rates, installation rates, leakage,[6] and hours of use,[7] among others), but these parameters were typically perceived as being more easily measured because they measure what *did* happen, not what *would* have happened, as rebound,[8] free ridership and spillover do.[9] This paper does not address these "other adjustments" in detail because we consider them to be distinct from the concepts of free ridership and spillover, which require the measurement of what did not happen. We do, however, return to the concept of adjusted gross savings as distinct from net savings in the conclusions and recommendations of this executive summary, and provide a more in-depth discussion in the full report.

Arguments for and against Measuring Net Savings

The third contextual issue affecting our research findings involves arguments for and against measuring net savings. The arguments in support of measuring net savings include the following:

- Establishes that the program brought about energy savings that would not have occurred in its absence
- Quantifies the amount of savings achieved as a result of the program in a specific time period
- Assesses the effectiveness of various program designs and whether the designs should be replicated, expanded, revised, or discontinued
- Ensures that ratepayer or taxpayer funds are being spent responsibly, and in a manner ensures that efficiency truly is the lowest-cost resource
- Uncovers fraudulent program implementation practices such as the claiming of savings from installations that occurred prior to any interaction with the program

[6] Leakage in the energy efficiency community refers to a measure incented by the program that is installed outside of the program's jurisdiction.. This is distinct from the use of the term leakage in the air regulation community. For this definition, see Horowitz (2009) in the Forum's *Glossary of Terms*.

[7] Not just total hours of use are important, but also by time of day and year, for purposes of estimating both demand impacts and emission impacts—the latter because power plants that are operating at different times have different emissions profiles.

[8] The Forum's *Glossary of Terms* defines the rebound effect as follows: "Also called Snap Back. A change in energy-using behavior that yields an increased level of service that is accompanied by an increase in energy use and occurs as a result of taking an energy efficiency action. The result of this effect is that the savings associated with the direct energy efficiency action is reduced by the resulting behavioral change."

[9] Measurement of these other adjustments is not without its challenges, even in the more precise metering and on-site verification studies. For example, participants who agree to take part in an onsite-verification may be more enthusiastic about the measure, leading to upward bias in estimates of installation rates and gross savings. Adjusted gross savings measurements that rely on self-reported installation and usage, including before and after participation, are subject to many of the same challenges as self-reported free ridership and spillover.

- Complies with regulatory requirements to report estimates of net savings along with gross savings

The arguments presented against measuring net savings can be summarized as follows:

- Focuses too heavily on narrowly defined metrics of individual program success or failure—especially free ridership—while deemphasizing other important impacts, such as non-energy benefits and behavioral effects, as well as portfolio- and policy-level impacts[10]
- Creates the impression that the estimates accurately represent the savings attributable to the program when, in reality, the methods are unreliable and often yield estimates that are biased[11] or lack validity such that the results may not accurately represent actual program-induced savings
- Requires expenditures of resources that are not in keeping with the importance of the estimates and their reliability or validity, thereby diverting resources from other planning, evaluation, and implementation activities that could yield greater benefits

It is worth noting that very few of the interviewees or literature sources we consulted argued that programs should not have to measure net savings; instead, the sources citing difficulties with the measurement of net savings called for de-emphasizing its importance or using ranges instead of point estimates to better reflect the uncertainty in the estimates. Free ridership is of particular concern to these commentators; they argue that the concept is often measured at the expense of other components of net program impacts such as spillover, non-energy impacts, or market transformation.

Measuring Net Impacts in Other Evaluation Fields

The project team performed a limited investigation into the question of how other fields of program evaluation address net impacts, and how often they are expected to measure them, if at all. If texts on program evaluation are a guide, it is considered a best practice in evaluation to document the extent to which programs have produced intended outcomes that would not have occurred in the absence of the programs. The evaluation literature also demonstrates that other evaluation fields face similar challenges as those facing the energy efficiency community, including having to assess components of net impacts that are similar to free ridership and spillover, as well as addressing issues of attribution in the face of multiple programs and messaging designed to produce the desired outcome. The degree to which programs actually

[10] By portfolio-level impacts, the commentators referred to measuring net savings for the entire suite of programs offered by a program administrator, sometimes with divisions between residential and non-residential programs. Regarding policy-level impacts, some—but by no means all—commentators believe that measuring goals such as job creation or emissions reductions should be based on gross estimates, including achievements that would still have occurred even without the program.

[11] While "bias" may not be *inherent* in the methods, most commentators asserted that, as currently implemented, most measurement methods systematically misrepresented the actual rates of free ridership and spillover. The commentators did not speak with one voice, however, on the direction of the bias for each concept. For example, some argued that free ridership is typically overestimated, while others said it tends to be underestimated.

undergo net impact evaluation, however, is less clear. Some programs never move beyond the level of evaluation known as "performance measurement," which is akin to tracking program data and reviewing it to make sure that it is achieving key milestones and outputs. Others measure gross impacts but do not take the additional step of documenting net impacts. Some programs, however, are required to undergo net impact evaluation, although this often occurs toward the end of the program or, if the program is on-going, only once every few years. If we were to generalize, it appears that programs that are larger and have more substantial evaluation budgets—particularly those in the fields of public health, social services, and education—are typically expected to demonstrate "net impacts," with the timing depending on the nature of the program and the requirements of the funding source.

Advantages and Disadvantages of Current Net Savings Methods

The challenges of measuring net savings are not new. Many of the conceptual and methodological critiques raised today echo those of twenty years ago, despite the fact that methods of measurements have greatly improved in the interim. The common methods employed to date have widely recognized flaws, as does the quality of data that serve as inputs to the methods. Although some methods of calculating net savings provide more statistically reliable and valid estimates than others, the attempt to measure net savings will always embody some degree of uncertainty precisely because it forces evaluators to estimate the counterfactual, something that never happened.[12] This forces us to confront the issue of construct validity, that is, the degree to which a concept measured through research matches the theoretical definition of the same concept. In other words, because we can never actually measure what did not happen, we will never be sure that our methods have accurately captured what a participant would have done absent the program. Evaluators have attempted to overcome this challenge by continuously modifying methods of measuring net savings, but most improvements end up being only incremental in nature because we can never rid ourselves of the counterfactual. In fact, it is this characteristic that leads some commentators, such as the independent system operators (ISOs) and some air regulators, to argue for basing reductions in energy demand and greenhouse gas emissions on adjusted gross savings and not net savings, based on the argument that the measurement of net is too imprecise to fulfill their directives to supply quantitative estimates.[13] The body of this paper summarizes the key methodologies currently used to assess net energy savings in the energy efficiency field, along with their strengths and weaknesses, and provides examples of studies relying on the various approaches (see Section 3, Table 3–1).

[12] The counterfactual enters into estimating adjusted gross savings as well. For example, when taking measurements (physical or self-reported) of energy use before and after the implementation of a measure, the evaluator assumes that the prior usage would have continued had the program not occurred. The key difference is that, for adjusted gross savings, one assumes the counterfactual, while in net savings estimation, one attempts to measure the counterfactual.

[13] Air regulators must demonstrate, for example, that measures are permanent, often limiting their ability to claim savings from behavioral and perhaps some market transformation programs if metering is not feasible.

Furthermore, data quality is a limiting factor for all methods; poor data quality can exacerbate concerns about reliability, validity, and bias. Currently, evaluators have consistent access only to program tracking databases. They often have access to participant billing data, but not always for non-participants; moreover, billing data are sometimes available only for a few recent years. Similar to program administrators, evaluators rarely have access to sales and shipment data at the state, regional, or national market level. This inconsistent access to high-quality data can force evaluators to rely on self-report methods or incomplete sales and shipment data to assess net savings. An alternative approach that others have begun to explore is taking a big picture modeling approach that uses macroeconomic data and attempts to identify the impact of efficiency programs using this top down approach. The approach is currently too new to have been explored in detail in this report.

Likewise, the inability to access sales data over time and by area for energy-efficient measures and equivalent standard-efficiency measures limits the ability of evaluators to assess the cumulative effects of market transformation programs. Moreover, the reliability and validity of predominantly qualitative approaches such as structured judgment (Delphi panels), historical tracing, and *modus operandi* methods may also suffer if low-quality data are used to inform the efforts.[14] Without higher-quality, representative data, methodological improvements to net savings estimates will necessarily be incremental. Big leaps in the reliability and validity of net savings estimates will come only when better data are made available that support the evaluation efforts. It is important to stress that the data themselves will not yield the net savings estimates, but that higher-quality data would mean improved inputs for the methods currently in use and perhaps would allow the energy efficiency community to explore the use of new methods that could take advantage of the better data.

Net Savings in Relation to Current and Evolving Policy Needs

Recent local, state, and federal legislation and policies targeting climate change, national energy independence, and economic stimulus have joined with more traditional energy efficiency program drivers to bring about expanded goals for energy savings, but also goals for important non-energy benefits such as greenhouse gas reductions and job creation, among other desired outcomes. These policies come not only with increased funds for implementing and evaluating energy efficiency programs, but also with increased scrutiny of program performance, as well as requirements for which outcomes are reported and in what manner. In some cases, energy or

[14] Delphi Panels provide experts with information on a topic, and each panelist makes a judgment on the topic and submits the information back to the evaluators. The evaluators compile the information and then resend them to the panel members, asking if they stand by their original judgments or if the different assessments of their peers have caused them to alter their judgments. The process can be repeated, but at least two rounds of judgment are required for a Delphi panel. Historical tracing and *modus operandi* approaches look for evidence in support—or refutation—of a program effect and draw conclusions based on the weight of evidence on whether a net program impact does or does not exist. These last two approaches are generally considered qualitative techniques because they usually rely on review of secondary literature, tracking database, and logic models; however, the evidence brought to bear could include primary quantitative analysis.

demand savings—be they gross or net—represent the primary outcome of interest (*e.g.* for the forward capacity markets), but in others, the calculation of energy savings is a critical step in the calculation of the targeted outcome, such as reductions in greenhouse gas emissions. We assessed the degree to which interviewees said that net savings approaches meet current and evolving policy needs.

Meeting Current Policy Needs

Commentators had mixed opinions as to whether net savings approaches meet current policy needs, as summarized below. Often, the same individual would argue that the current approaches meet some policy needs but not others. Generally, energy regulators were more likely to state that the approaches meet current policy needs, while program administrators and net savings experts voiced greater skepticism on this topic.

Net Savings Approaches Do Meet Current Policy Needs. Those saying that current approaches do meet current policy needs asserted that net savings estimates provide a critical piece of information about how well the program is performing. Net savings estimates help program administrators, planners, and regulators understand the strengths and weaknesses of the program design and decide when to stay the course, revise a program, or discontinue it. Those holding this opinion argued that, although other program impacts should be evaluated, net savings should remain one of the most important considerations, largely to ensure that ratepayer and taxpayer funds are spent wisely to achieve the greatest return on investment in terms of energy savings.

Net Savings Approaches Do Not Meet Current Policy Needs. Commentators offered two explanations for why they did not believe that approaches to net savings meet current policy needs. The first explanation rested on skepticism regarding the quality of existing methods and, therefore, the results. Approaches to net savings could meet current policy needs, many of these individuals argued, if they could more accurately estimate free ridership and spillover, as well as cumulative effects and market effects.[15] The second explanation was that net savings is too narrow a focus and fails to recognize the broader context in which current programs (and future ones too) operate. These people argued that programs should be evaluated in terms of how well they engage the customer and should involve measures like extent of behavioral change and market transformation that could be thought of as a very broad and inclusive definition of spillover. It is important to note that this group thought that the current policy focus on net savings—more specifically free ridership—impairs planning and innovation that will allow programs to meet evolving policy needs.

Meeting Evolving Policy Needs

We asked interviewees to discuss the evolving policy needs regarding net savings, particularly those focused on expanded energy savings goals, reduced greenhouse gas emissions, and jobs

[15] This would include market effects not only of explicit market transformation programs, but also resource acquisition programs that potentially have market-level impacts.

creation. Policies of interest included energy savings goals at the state level, the Regional Greenhouse Gas Initiative (RGGI) in the Northeast[16], the federal American Recovery and Reinvestment Act (ARRA) and similar state stimulus funds slotted for energy efficiency programs and green jobs. The take-away message can be summarized as follows: 1) Things are changing fast, 2) Programs will need to adopt new design and implementation approaches to achieve the ambitious goals that have been set, and 3) Program evaluation will need to make significant adjustments in response to these changes, but 4) We are not quite sure what these changes will entail or what will be needed, so we are uncertain about what direction to take. These circumstances contributed to differences regarding the degree to which interviewees said that current approaches to net savings would meet evolving policy needs.

These differences were perhaps most notable regarding air regulation. While RGGI has provided funds to existing energy efficiency programs (or new programs very similar to existing ones), thereby supporting current evaluation approaches, there is the potential for a federal greenhouse gas regulatory program to exist in the future. Multiple commentators believe it makes sense for the energy efficiency community to plan for this eventuality, not by designing new programs, but by working with the air regulation community to develop evaluation and reporting approaches that support both communities. Like their energy efficiency counterparts, air regulators would like reassurance that emissions estimates resulting from efficiency programs are valid and measurable. Air regulators and many commentators in the efficiency community noted that savings estimates would require new metrics that tie energy efficiency impacts to emissions reductions at the power plant level, most likely by time of day and time of year. Yet disagreement remains within and between both the air regulation and energy efficiency communities about the degree to which current energy efficiency practices do or can meet the air regulatory needs. A pivotal issue relates to whether the air regulation community will be required to base savings on adjusted gross or net savings. A related concern is which methods of estimating energy savings will be acceptable for "proving" that emissions reductions actually occurred.[17] One concern with "provable" is that the ambitious goals for energy savings and emissions reductions currently in place in multiple jurisdictions in the Northeast are expected to be achieved through multiple programs, including behavioral ones and others for which it may be challenging to measure gross savings, let alone net savings. This raises the question of whether such programs will be penalized or their savings not claimed because they cannot prove their more difficult-to-measure savings. Another issue is how to define baseline in projections of emissions, particularly whether those baseline projections assume the emissions reductions

[16] RGGI currently does not have regulatory overview power, but it is possible, and some would say likely, that there will be directives at the federal level that will set forth directives concerning how to treat energy savings in estimating emissions reductions.

[17] Calls for providing quantifiable savings lead some in the air regulation community to argue for using adjusted gross savings, while others point to the concept of additionality as suggesting the need for net savings. Further complicating matters is the belief among some that additionality simply means not double counting reductions while others assert that it refers to measuring net reductions.

resulting from energy efficiency programs or if such reductions will be credited to the programs as achievements beyond the baseline.

At this time, there is no policy requiring the development of energy saving estimation methods within the energy efficiency community that also meet the needs of the air regulation community. However, some commentators said that the "writing is on the wall" that such policy is forthcoming. These individuals asserted that the needs of the energy efficiency and air regulation communities would best be met if these methods were developed in partnership, rather than having the air regulators decide the methods and the energy efficiency community react to them. Likewise, system planners would need to be a part of such conversations, as emissions reductions depend on which power plants operate at the time savings are achieved. Key issues to address will include not only deciding whether to report adjusted gross or net savings, but also how to incorporate the concept of additionality, and ensuring that methods meet the requirements set forth in State Implementation Plans (SIPs).

Attribution in the Face of Multiple Programs and Policies

As mentioned earlier, attribution refers to the practice of determining what impacts are *caused* by a specific program or portfolio. The estimation of net energy savings—what portion of gross savings a specific program has brought about—always involves determining attribution, but attribution analysis does not always produce estimates of net savings. In other words, to many commentators estimating free ridership and spillover *can* be a part of attribution analysis, but it need not be. To others, attribution is the estimation of net savings. Increasingly, the energy efficiency community is focusing a great deal of attention on one aspect of attribution: isolating the impact of the program from the impacts of other programs, messages, and contexts that could also influence the same or similar energy saving actions. This focus reflects the perception of some commentators that the complex web of programs, messages, and influences that encourage individuals to adopt energy efficiency measures makes the task of estimating net savings more difficult than before—ultimately, they argue, reducing the accuracy of recent net savings estimates.

To some extent, the task of disentangling the effects of a particular program from other influences encouraging the same action remains an important one for estimating net savings. Yet it is important not to lose sight of the overall goal of reducing energy use by focusing too narrowly on which program or source of funds gets the credit, because there could be synergistic effects among programs; the whole may be greater than the sum of the parts. There is a danger that the effectiveness of programs could be reduced if the energy efficiency community focuses program efforts only on what is most provable, not what is most effective. In fact, some commentators argued that estimation of net savings at the programmatic level should be dropped in favor of measuring savings (and emissions reductions) for the entire portfolio of programs offered by a program administrator or within a given jurisdiction or region.

The challenges of attribution in the general sense of sorting out causal connections that have multiple roots, some psychological and others practical, is addressed in more detail in the full report. In brief, some of the critical psychological issues involved include the following:

- After having taken an action, individuals begin to see themselves as "just that sort of person" and tend to attribute their past actions to this characteristic of themselves
- Intention does not always equal action; therefore many individuals may intend to take the energy-saving action but not act until exposed to the program—or, more accurately, its incentive. The question becomes how much credit to give to the program versus the other potential reasons for action that existed but had not induced the behavior.

Interviewees voiced a great deal of concern about attribution, with some seeing it as one of the most critical issues facing net savings evaluation. Even so, neither the interviewees nor the literature offered much guidance on how to address attribution. One suggestion involves turning to predominantly qualitative techniques such as historical tracing or *modus operandi* that draw on a wide variety of sources to search for evidence supporting the hypothesis that the program caused the savings while simultaneously ruling out the possibility that other factors caused the savings. The researcher behaves similarly to a detective, building a body of evidence in support of program causation by ruling out all other possible causes of the targeted action, or concludes that the evidence does not support causation by the program. Usually the evidence involves reviewing secondary sources, tracking databases, and logic models, but it could involve primary qualitative research such as interviewing or quantitative analyses of data as well. This approach works best when the program planners and administrators have developed a carefully constructed program theory and logic model showing how the causal relationships are *supposed* to work, making it easier to search for the necessary connections to demonstrate that they *have* worked. Other suggested approaches include probabilistic methods such as statistical modeling, assigning impacts in proportion to leveraged resources going into the same program or effort, which applies only to some situations,[18] and the use of structured judgment approaches such as Delphi panels.

The Issue of Consistency in Net Savings Definitions and Methods

The project team explored interviewees' opinions about the possibility of adopting consistent definitions of and methods for measuring net savings throughout the Northeast. In some ways, the definition of net savings guides the methods; one must know which components are included in order to use each method. However, numerous methods can often provide the information required by the same definition or conceptualization of net savings, although the resulting estimates and their validity may differ. For this reason, net savings methods differ a great deal, even among jurisdictions using the same definition—or even within a single jurisdiction—let

[18] If one funding source provided 35% of funds, another 15% and a third 50%, the first would get credit for 35% of the savings, the second 15%, and the third 50%.

alone across jurisdictions with different definitions of net savings. Hence, methods for measuring net savings differ greatly even under situations in which the same definition is used.

Consistent Definitions of Net Savings

Most interviewees supported the idea of consistent definitions of net savings. The main reason the Northeast Region currently does not have consistent definitions, some argued, is that there has been no compelling public policy driver, such as legislation or interstate agreements, to stimulate the development of a consistent definition. Instead, jurisdictions have largely taken the path of least resistance by developing their own definitions, and many would be reluctant to let go of their definitions without some compelling reason to do so.

Consistent Methods for Measuring Net Savings

Opinions diverged to a greater degree regarding consistent methods for measuring net savings. Most respondents recognized some benefits of adopting consistent methods for measuring net savings, particularly in light of increased regional cooperation regarding greenhouse gas emissions, and in order to facilitate evaluation and reporting for programs and program administrators operating across jurisdictions. However, individuals voiced concern about overly prescriptive approaches that may not reflect the diverse situations in which programs operate, may not be feasible for programs with limited resources, could stifle creativity, and would draw attention from what some view as more important priorities, such as determining how to attribute program impacts in the age of multiple program and factors affecting the adoption of energy efficiency measures and behaviors. Others, while advocating consistency, also argued that a consistent approach should not include current methods because they are too fraught with reliability, validity, and bias concerns.

We also noted some patterns in opinions on net methods, although individuals may deviate from the pattern identified for their stakeholder groups. Air regulators were the strongest advocates for consistency in estimation methods, regardless of whether they wanted adjusted gross or net savings estimates.[19] Net savings experts and program administrators—including some working in numerous jurisdictions—generally liked the idea of consistent methods for net savings, but often raised concerns about doing so in practice, particularly given skepticism about the reliability and validity of methods and their ability to meet current or evolving policy needs. Energy regulators and their representatives voiced the greatest skepticism regarding the promotion and adoption of consistent methods, supporting their positions by pointing to the evolution of the diverse methods currently in use—namely that older methods have been found to be lacking or did not adequately represent the nature of particular programs or populations.

Overcoming Barriers to Consistency in Net Savings Definitions and Methods

Getting all of these relevant parties to agree on the details of a consistent approach could be difficult given the lack of a policy driver to stimulate development of consistent approaches. It is

[19] As mentioned above, air regulators did not speak with one voice on the issue of adjusted gross vs. net savings.

the opinion of the project team that most of the concerns raised about consistent approaches to net savings could be alleviated through the adoption of consistent yet flexible *guidelines* that avoid the rigidities that could stifle creativity or that could fail to address unique program circumstances.[20] However, the Forum must decide if it is indeed the goal of its members and the regions as a whole to adopt consistent approaches before devoting resources to the development of such guidelines for a consistent approach for measuring net savings. Before developing such guidelines, then, the Forum and its members would need a *framework* for how to achieve consistency.

Conclusions, Next Steps, Recommendations, and Future Research

The research presented in this scoping paper demonstrates that the definition, measurement, and uses of net savings have long challenged the energy efficiency community and continue to do so today. Some evidence suggests that these challenges have recently been exacerbated by an increase in regional cooperation on energy efficiency and climate change programs and policies, an expansion of the audiences for energy savings estimates, and a proliferation of influences (*e.g.,* programs, media coverage, and economic policies) affecting energy efficient action. Together, these persistent and growing complications lead to confusion and disagreement over such issues as the best way to demonstrate that a program *caused* savings to occur, whether the air regulation community will need adjusted gross or net energy savings to estimate emissions reductions, and whether the Northeast should pursue consistent definitions and measurement approaches to net savings.

Despite the continued confusion and debates, our research does support the continued use of net savings estimates for certain specific purposes, as follows:

- *Assessing the degree to which programs bring about a reduction in energy usage and demand*. While this is a controversial use of net savings estimates, the research supports the continued use of net savings as *one* of numerous measures that should be given serious consideration in the assessment of program success, at least until a suitable alternative is developed beyond gross or adjusted gross savings that recognizes that some energy and demand savings would have happened without the program.
- *Uncovering fraudulent program implementation practices*. Cases have been documented in which program implementers have claimed savings from activities that they clearly did not influence, including installations that occurred prior to any interaction with the program and random downward fluctuations in energy use in excess of what accrued to any program activity

[20] For example, if those guidelines expected each program administrator to adopt a transparent and common approach for all their programs, at least reliability, validity and bias would be consistent for each program administrator year to year.

- *Gaining insight into how the market is changing and transforming over time* by tracking net savings across program years and determining the extent to which free ridership and spillover rates and net-to-gross ratios have changed over the period.
- *Understanding better how the market responds to the program* and using the information to inform modifications to program design, including measure eligibility and targeted marketing. Later, these program modifications would again be subject to net savings evaluation, in an adaptive management process.

Recommendations and Research Needs

In addition to these four uses, the research also points to a series of recommendations and research needs related to pursuing consistent approaches to defining and measuring energy savings—net and gross—throughout the Northeast, securing higher quality data, determining program attribution, and aligning the research needs and activities of the energy efficiency community with those of the air regulation community. We present each recommendation below, and expand on them in the main body of the report.

Recommendation 1: The Forum should lead the process of developing a consistent definition of *adjusted gross savings* in the Northeast Region. The research activities conducted for this scoping paper revealed that different jurisdictions in the Northeast Region and the nation vary in the degree to which they measure not only free ridership and spillover but also such parameters as rebound, leakage, hours of use, installation rates, and persistence rates, among others, and whether these additional parameters are considered to be adjustments to gross savings or elements of net savings. It is the *opinion* of the project team that all of these concepts *except free ridership, spillover, and rebound* constitute adjustments to gross savings as they measure what actually *has* happened, as opposed to free ridership, spillover, and rebound which involve the measurement of what *would have* happened in the absence of the program. Ultimately, the Forum will have to decide if it agrees with our recommendation and, if so, which of the possible adjustments will be included in its definition of adjusted gross savings.

Related Research Needs for Recommendation 1:

- Identifying the adjustments that are currently submitted to the ISOs for reporting in support of the forward capacity markets.
- Examining the components that factor into realization rates, which also are not consistently defined from program to program, jurisdiction to jurisdiction, or evaluation to evaluation.
- Evaluating the degree to which our recommended components for adjusted gross savings meet the needs and requirements of the energy efficiency community in the Northeast, given that our efforts focused on net savings and not adjusted gross savings.

Recommendation 2: After developing a consistent definition of adjusted gross savings, the Forum should develop a definition of *net savings* for the Northeast Region. This definition would be in relation to—but separate from—that of adjusted gross savings. The Forum would

need to consider which concepts should be included in this definition, particularly focusing on the treatment of free ridership and spillover.

Related Research Needs for Recommendation 2:

- Enumerating in detail the similarities and differences between current definitions of net savings in use throughout the Northeast.
- Understanding the reasons for differences in definitions of net savings and potential ways to bring the definitions into alignment while meeting the needs of the diverse energy efficiency community in the Region.

Recommendation 3: The Forum and its members and allies should consider advocating for legal requirements for manufacturers, retailers, and distributors to provide national sales and shipment data for key equipment and products—reported by size and efficiency, ideally at the county, or at least state, level. Legally requiring reporting of these data may provide the only avenue for accurate estimation of net savings from upstream market transformation programs—including capturing the savings resulting from cumulative program activity from prior years. Higher-quality data will not only improve estimation from quantitative approaches, but they could also provide more dependable information for use in qualitative techniques such as structured judgment (Delphi panels), weight-of-evidence, and historical tracing approaches. Note that this recommendation does not require further research *per se*, but instead would focus on advocacy, perhaps accompanied by a research project to facilitate the advocacy and develop supporting materials on the part of the Forum and its members.

Recommendation 4: Program administrators should keep records of program activity by year (*e.g.*, budgets, measure covered, participants served, previous estimates of net savings, *etc*.), including in any possible comparison areas. Such data will assist in the development of net savings estimates by allowing for the estimation of cumulative and spillover effects over time, thereby providing a more complete assessment of program impacts. Although this recommendation must largely be implemented by program administrators, the Forum could assist program administrators in the development of data collection and storage tools, perhaps even developing a database that contains such information for easy reference and comparison across areas and years. Currently, the Consortium for Energy Efficiency keeps track of program data for members who respond to their annual request for information, and their experiences should certainly be taken into account.

Recommendation 5: The Forum should also clarify the definition of *attribution* and the degree to which programs must differentiate the impact of their activities from the impacts of other programs and efforts designed to bring about the same or similar actions or outcomes, and from contextual factors such as energy prices and the economy. The term "attribution" is consistently applied to capture the impacts *caused* by program activity, but is increasingly being used to refer to separating the effects of *this* program from the effects of all the other programs or pro-efficiency messages to which participants have been exposed, and

from the effects of contextual factors such as energy prices and the economy, during a specific period of time. While it may be important to assess the impacts brought about by the program, the energy efficiency community should be careful not to lose sight of the goal of reducing energy use by focusing too narrowly on which program gets the credit, particularly as various programs and influences may have synergistic effects such that the whole of their combined impact is more than the sum of their individual impacts.

Research Needs for Recommendation 5:

- Delineating the percentage of participants that are "repeat players"—that is, those taking part in the same program in different funding cycles or taking part in multiple programs offered by the same program administrator in the same or different funding cycles.
- Comparing the total savings of repeat players to the aggregation of savings among comparable individual program participants (*i.e.,* the savings from Participant A from program X plus the savings from Participant B from Program Y, and so forth).
- Identifying the variety of potential influences outside of program activity that may affect program participation and assessing whether direct causal links can be established between those other potential influences and energy saving actions.

Recommendation 6: The Forum should encourage the energy efficiency community—particularly energy regulators—to expand its assessment of program success from a focus on net savings, particularly free ridership, to the inclusion of additional factors that may more accurately capture the full range of program impacts, including non-energy impacts such as jobs, improved health, and increased productivity. Net savings is a valuable measure of the amount of savings the program has achieved that would not have happened otherwise. But, as this scoping paper has documented, net savings measures are not infallible, nor do they fully capture the wide range of impacts that may result from energy efficiency program activity. Therefore, while we recommend the continued consideration of net savings estimates, we also suggest that additional impacts be taken into account when assessing programs and determining any reward or penalties to be paid.

Research Needs for Recommendation 6:

- Exploring approaches that best capture the cumulative—that is, long-term—effects of market transformation programs on the targeted markets.
- Identifying the impacts beyond energy savings that are most crucial to determining program performance and ensuring that ratepayer and taxpayer funds are being used responsibly
- Developing a prioritization scheme that may assist energy regulators in expanding their assessment of program performance beyond net energy savings and its components

Recommendation 7: The Forum should decide if it supports the development of consistent methodological approaches to estimating net savings for the Northeast, and, if so, take the actions necessary to develop regional guidelines for consistent methods. This

recommendation involves three different steps. First, the Forum should discuss the reasons for and against developing consistent approaches to net savings estimation and decide whether consistency should be pursued in the Northeast. Second, should the Forum recommend in favor of consistency in methods for estimating net savings, it should begin working toward consistency by first developing a framework for how to achieve this consistency. The framework would set forth initial suggestions on how program administrators could begin moving toward consistency. The framework could also identify some of the key considerations that program administrators and evaluators should take into account when choosing from among the various methods for measuring net savings. This framework would serve as an interim step toward consistency while the Forum waits for the results of research projects that would be needed before embarking on the third and final step: the development of guidelines for consistent methodological approaches to net savings. The guidelines would provide more explicit recommendations concerning which methodological approaches to use in specific situations. The guidelines should also allow for the introduction of new methods, with a process or criteria for establishing their reliability, validity, and rigor. Although flexible, the guidelines should avoid the trap of "anything goes" by specifically identifying the best approaches to be pursued given varying levels of resources and by identifying those approaches that should be avoided except in very limited circumstances.

Research Needs for Recommendation 7:

- Understanding the range of methods available for estimating net savings, their strengths and weaknesses, and their ability to provide the types of estimates needed by the energy efficiency community (and perhaps the air regulation community) to ensure that any guidelines for consistent methodological approaches to net savings focus on the best net impact evaluation practices available.
- Exploring the existing academic and energy efficiency research on the psychological and sociological processes that influence estimates of net energy savings in an effort to understand more fully the ways in which these processes may affect how participants respond to self-report questions about their past actions and likely behavior in the absence of the program.
- Examining the potential of macroeconomic approaches for estimating the impact of program activity on net energy savings, recognizing that such approaches are new and unproven at this time.

Recommendation 8: The Forum should facilitate the development of a working group comprising members of the energy efficiency community, the system planning community, and the air regulation community with the ultimate goal of developing approaches to measuring energy savings and resultant reductions in greenhouse gas emissions in a manner that is mutually acceptable to and feasible for all three communities. This recommendation rests on the assumption that the federal government will likely adopt guidelines on how to account for emissions reductions resulting from energy savings achieved through efficiency programs. It further assumes that the energy efficiency community would prefer to be

an active part of the development of these guidelines rather than responding to guidelines set forth by the air regulation community. Given these assumptions, such a working group would have three tasks. The first task would be to develop a mutual understanding of the terminology used in each field and to identify compatible concepts across both fields. Where concepts are not compatible, the working group would seek to bring them into closer alignment to the extent possible while still meeting the needs of the two communities. The second task would be to review methodologies for estimating baseline emissions reductions, gross energy savings, and net energy savings, making certain that the baselines for emissions reductions do not include projections of energy savings expected to come from efficiency programs and that the methodologies for estimating energy savings fulfill the SIP verification processes. In the final task, the working group would deliver a set of recommendations to the EPA, and state energy and air regulation agencies on how best to account for energy savings from emissions reductions in a manner that is mutually agreeable to both the energy efficiency and air regulation communities.

Research Needs for Recommendation 8:

- Examining the possible methods or approaches for translating energy efficiency impacts into measureable pollution reductions from power plants by time of day and time of year.
- Exploring possible approaches to measuring *net energy savings* that would meet the requirements for reliability and precision set forth in SIPs. This may involve the use of technology such as advanced metering (if a method to translate the information to net savings were developed) or the increased use of randomized controlled experiments in estimating net savings.

The project team believes that the energy efficiency community—and probably the air regulation community as well—should acknowledge the fact that energy efficiency and emissions reductions programs are not wholly responsible for all the savings and emissions reductions that may be achieved through the adoption of the devices or behaviors promoted by the programs. The estimation of net impacts—be they savings, emissions reductions, jobs created, water saved, and so on—is a way of making this acknowledgement. Currently, the measurement of net impacts occurs at the measure or program level, with the results sometimes being aggregated to a group of programs or to the entire portfolio offered by a program administrator. However, we recognize that demonstrating attribution has become increasingly challenging in light of numerous programs and factors that may influence the targeted behaviors Moreover, jurisdictions at the municipal to state levels—and perhaps one day at the federal level—have set ambitious goals to reduce energy use and greenhouse gas emissions. Given the uncertainty inherent in measuring the counterfactual of net savings, some members of the energy efficiency community are beginning to ask if we need a radical rethinking of how we go about estimating net impacts in general and net savings in particular. It is beyond the scope of this paper to describe what such a radical rethinking of impact evaluation methods might entail, but we do offer a few interim recommended alternatives to *measuring* net savings that nevertheless still

recognize that program activity does not necessarily yield all the measured gross savings. This brings us to the final recommendation.

Recommendation 9: The Forum should consider the potential of using a deemed or negotiated net savings approach for crediting energy savings—or emissions reductions—to a program or portfolio. In such an approach, program administrators and regulators would draw on available evidence and additional indicators of program impacts to help decide on the percentage of gross savings that can be claimed by a program or portfolio after it has carefully demonstrated that the program activities have strongly contributed to the desired outcomes. The credibility of the negotiated net savings figure would depend on the type, amount, and quality of the information informing it; such information could include program tracking data, adjusted gross savings estimates, net savings estimates derived from periodic research (possibly multiple approaches), the sales/shipment data mentioned in Recommendation 3, market research to assess the state of the market, energy intensity by sector over time, and more. The percentage could be developed through a Delphi panel or using other structured expert judgment approaches. Importantly, the approach should be negotiated and agreed on beforehand; it is even possible—as with the recent decision in Arizona to credit up to one-third of the savings from utility C&S efforts toward its 2020 Energy Efficiency Standards (EES) target, provided they make a credible effort—to decide on the number beforehand. The conclusion would still be subjective, and its precision—how close it is to the real value of net savings—would be uncertain because achievement of the goal could be questioned on the basis of the counterfactual as well as factors such as economic recessions, reduction in certain types of manufacturing, changes in energy prices, and so forth. Even so, the negotiated agreement approach avoids depending entirely on controversial measurements of the counterfactual of net savings.

To conclude, the definition and measurement of net energy savings remains a controversial issue in the energy efficiency community. The controversy surrounding the issue presents challenges to the task of achieving consistency in definitions and measurement of net savings throughout the Northeast. Some commentators do not believe consistency is needed or desirable, while others perceive the appeal of consistency but do not believe it can be achieved in practice. Other commentators, however, support the idea of increased regional consistency in definitions and measurement of net savings, citing increased cooperation in reducing energy use and greenhouse gas emissions across state lines and the fact that many program administrators operate in multiple jurisdictions. This paper has sought to clarify some of the issues surrounding net savings definitions and measurement and the pursuit of regional consistency. The recommended action and research needs follow directly from the observations identified in the literature, by interviewees, and by various Forum and Steering Committee members. It is our hope that the findings and recommendations will help the Forum—and the regional energy efficiency community—come to some agreement regarding the definition and measurement of adjusted gross savings and net savings, whether to pursue consistency in the near future, and the degree to which the needs of the air regulation community may be a driver of both.

1 Introduction

NMR Group, Inc. and its subcontractor Research Into Action, Inc. (the project team) are pleased to submit this scoping paper on Net Savings to Northeast Energy Efficiency Partnerships, Inc. (NEEP) and its Evaluation, Measurement, and Verification (EM&V) Forum (hereafter EMV Forum or just Forum).

1.1 Background

The goal of the project is to produce a scoping paper that informs and supports the following EMV Forum goals: understanding the reasons for measuring net savings; increasing the consistency and quality in EM&V practices with respect to defining and estimating energy efficiency program net savings within the Northeast Region; and identifying the needs of stakeholders (*e.g.*, air regulators[21]) who might (or might not) use net savings estimates to measure progress toward non-energy goals. The specific project objectives are as follows:

- To identify the key audiences that rely on net savings to meet policy and program needs
- To highlight the reasons for measuring net savings and the situations in which such measurement is and is not appropriate
- To document the issues and challenges associated with existing approaches to defining and measuring net savings
- To propose recommendations for next steps needed to make progress toward the goal of increasing the consistency and quality of definitions and methods related to net savings

For over 20 years, the energy efficiency community has given significant attention to the issue of how to define and measure *net savings*—that is, the amount of energy savings that is causally attributable to an energy efficiency program or programs. Despite these efforts, conceptual debates and methodological challenges persist within the energy efficiency community (Saxonis 1991a, 2007a, 2010; TecMarket Works 2004; Skumatz, Khawaja, and Colby 2009; Messenger, Bharvirkar, Golemboski, Goldman, and Schiller 2010). Although a seemingly straightforward concept, the counterfactual of net savings confronts us with three broad categories of problems:

- *Problems with construct validity*: the conceptualization of what would have happened in the absence of the program (the counterfactual) is limited by the necessity of circumscribing the scope to a certain set of activities over a certain period of time. This problem feeds the lack of consistency found in definitions and measurement approaches to net savings.
- *Problems with measurement approaches*: all methods employed to date have widely recognized flaws (although some flaws are more problematic than others) as do the quality of data that serve as inputs to the methods. Again, this problem exacerbates

[21] For the purposes of this papers, "air regulators" refers to federal and state environmental regulators focused on greenhouse gas emissions.

differences in measurement approaches, as evaluators attempt to improve methods. As it is, most improvements end up being only incremental in nature.

- *Problems with the application of net savings estimates*: regulators make decisions based on net savings results that affect program administrators financially. On the basis of net savings estimates, regulators or program administrators may alter or suspend programs, often without recognition of the problems with construct validity and measurement. For example, the Massachusetts ENERGY STAR® Appliances Program adjusted its program to provide incentives on only Consortium for Energy Efficiency Tier 3 (highest efficiency) clothes washers after evaluators estimated low net-to-gross ratios for a program that incented any ENERGY STAR®-qualifying model. Eventually, the program removed all incentives for clothes washers. The importance assigned to net savings estimates makes them subject to scrutiny and a "hot button" topic in the energy efficiency community, where there are widely varying opinions on whether available approaches meet current and evolving policy needs and the degree to which the Northeast should pursue consistent approaches across jurisdictions.

It is also the case that the stakeholders who are concerned with and will likely use net savings estimates—what we will call the "audience" for net savings—has recently diversified in response to increased funding and expanded goals and objectives for energy efficiency programs (In-depth interviews; Saxonis 2007a). These new audiences extend the scope of the debates and challenges, and subject net savings estimates to further scrutiny as they have their own ideas about whether and how net savings should be estimated and the degree to which definitions and methods for estimating net savings across different regions and jurisdictions should be consistent. At times, the ideas of the new audiences may contradict those of the energy efficiency community.[22] The energy efficiency community, therefore, will need a clear understanding of the issues and what is at stake in order to meet the program delivery, energy savings, and emissions reduction goals laid out for them. This scoping paper seeks to add to this understanding of the pertinent issues by addressing definitions of net savings, measurement concerns, and issues involving consistency in approaches to net savings.

Throughout the course of this project, the project team frequently interacted with NEEP staff and Forum members to clarify the objectives as the project progressed. While the scoping paper still addresses each of the objectives outlined in the work plan, it became clear that focusing on the intersection of measurement approaches, policy needs, and consistent definitions and methods would provide the greatest benefit to the Forum. Other authors have extensively documented the different methods of measuring net savings and their strengths and weaknesses (*e.g.* Saxonis 1991a; Goldberg, Bloch, Prahl, Sumi, Ward, Winch, and Talerico 2006; Skumatz *et al.* 2009; Rosenberg and Hoefgen 2009; Hoefgen 2010). For this reason, we focus our discussion of measurement approaches and their strengths and weaknesses on the concerns about current and

[22] For example, variations in program design across jurisdictions may conflict with regional calls for consistent measurement of net savings.

evolving policy and regional consistency with which the Forum is grappling, including attribution in the face of multiple programs and messages, addressing the needs of diverse audiences for net savings, meeting the expanded energy savings and non-energy benefit goals of efficiency programs, and the advantages, disadvantages, and challenges of formulating consistent definitions and methods for net savings across the Northeast Region.

1.2 Methods

To meet the objectives of this paper, the project team performed an extensive literature review of approximately 100 articles, papers, presentations, and book chapters from within the field of energy efficiency as well as areas outside the field, including public health, education, environmental studies, and program evaluation methodology. We also interviewed 12 experts on the administration, implementation, and evaluation of energy efficiency programs for energy regulation and air regulation to gain their insights into issues related to energy savings in general and net savings in particular.[23] Finally, we incorporated insights shared by Forum and subcommittee members after they reviewed the initial draft of this document. The remainder of this section describes the research methods in more detail.

1.2.1 Literature Review

The literature review serves two important purposes. First, it provides a context—over time, across locations, and beyond energy efficiency—for issues related to program impacts generally and net savings in particular. Second, the process of conducting the literature review informed the in-depth interviews by pointing to issues to address with the experts.

The literature included in the review is primarily a convenience sample comprising resources previously known to the project team and/or Forum members. We also consulted some of the works cited in these initial resources that appeared to have relevancy for this scoping paper. In addition, we identified a few additional sources through internet searches of academic journal indices (*e.g.* ERIC, Pro-Quest) and the website for the American Evaluation Association (AEA), from which members can access online versions of AEA publications.[24] Finally, the project team and NEEP staff attended the American Council for an Energy-Efficient Economy's Summer Study in August 2010; certain papers presented at the Summer Study had relevancy for this scoping paper, and we incorporated those findings into this revised draft. The sources reviewed here are not necessarily representative of the entire body of literature on program evaluation within or outside of energy efficiency, but we believe they are sufficiently comprehensive for our purposes. They touch on the key topics and issues pertinent to our focus on net savings, address

[23] The independent system operators (ISOs) in the Northeast have decided to accept estimates of adjusted gross savings for the Forward Capacity Markets. For this reason, we did not include system planners among our interviewees. We recognize that their perspectives and needs may be relevant to consider in the future—for example, in the context of incorporating energy efficiency into system planning forecasts.

[24] The resources we reviewed included conference, scoping, or white papers, presentations developed primarily for energy-efficiency professionals (*e.g.*, program planners, implementers, evaluators, and regulators), final reports of actual impact evaluations, and articles published in peer-reviewed academic journals.

how researchers in other fields deal with analogous issues involved in measuring the impact of programs, and include literature on program evaluation theory and practice in general.

When reading the various sources, the team tracked what each contributed to our knowledge of the following five key topics:

1. How to define gross savings and net savings, free ridership, and spillover
2. Advantages and disadvantages of measuring net savings; situations in which it is appropriate or inappropriate to do so
3. Issues and challenges associated with estimating net savings, generally and related to specific concepts or methodologies
4. Key audiences for net savings estimates
5. Recommendations for next steps toward resolving issues related to net savings[25]

For each of these five topics, we documented the various themes and perspectives represented in the readings. We paid particular attention to the intersection between net savings and the policy and regulatory needs that energy efficiency programs are increasingly being asked to meet.

1.2.2 In-depth Interviews

While the literature review provided the team with an understanding of the dominant concerns, issues, and viewpoints related to net savings, the in-depth interviews allowed us to focus more specifically on the issues of greatest concern to the Forum that have not been adequately addressed in the existing literature—specifically the emerging issues of the expanding audiences for net savings estimates as well as current and potential uses to which the estimates may be applied.

Together, the Forum and project team identified potential interviewees based on their knowledge of energy savings estimation methods and policies or their familiarity with energy and/or air regulation and the needs of regulators. After limiting the focus of the interviews to current and emerging policy needs and shifting audiences, we targeted potential interviewees with knowledge of these issues who collectively represented diverse perspectives across the region and even the nation. We do not identify the interviewees in this document, as we assured them that their responses would not be associated with their names; given the close-knit nature of the community, we could not guarantee that readers of this report would not be able to identify the authors of particular responses if they knew who was interviewed.

The project team developed the interview guide in cooperation with the Forum (see Appendix A). The guide includes a few initial questions about how respondents define and conceptualize net savings, what they perceive as the appropriate roles (if any) for net savings in evaluation, and the strengths and weaknesses of current approaches to estimating net savings. The guide also addresses the more critical issues of audiences for net savings estimates, the role of net savings

[25] These include recommendations from articles written in the past that the evaluation community actually implemented after the publication of the article.

in allocating resources for programs and initiatives to reduce energy consumption and emissions, current and evolving policy needs in regard to net savings, and regional consistency in definitions and methods for net savings.

We limited our interviews to no more than 30 minutes unless the respondent agreed to continue beyond that length of time. To facilitate the interviews, we provided each interviewee with a copy of the guide prior to the interview. Interviewers did not ask the interviewees every question, nor did they ask the questions in the exact order listed in the guide. Instead, we tailored each interview to the expertise and interests of the respondent, always giving the respondent the opportunity to address topics related to net savings that were not in the guide.

The project team had originally planned to use the same interview guide for air regulators as for the energy-efficiency experts on our list. However, after reviewing a draft interview guide, air regulators on the Forum or Steering Committee commented that most of their colleagues had only a general understanding of net savings and would not be able to respond to many of our questions. Based on our conversations with Forum members, we developed a shorter guide for air regulators focusing specifically on their needs *vis-à-vis* energy savings reporting and methodology at a regional level. Appendix A includes this air regulator interview guide.

1.2.3 Forum Feedback

After the project team submitted a draft report to the Forum, Forum members provided feedback on the report and also shared their own opinions on the topics addressed in the study. The draft report seems to have helped Forum members on the project subcommittee to solidify some of their own thinking on the topics, and they weighed in with some helpful insights and observations. For this reason, NEEP directed the project team to treat such comments as data for inclusion in the study. Although many of the comments have been circulated to all Forum members on the project subcommittee, we still treat the material as confidential by not identifying the individual associated with the observation, as this paper will be shared beyond the Forum.

1.3 Important Terms and Concepts

One of the primary purposes of this scoping paper is to explore how net savings and related terms are defined and conceptualized in the practice. While these definitions and conceptualizations vary throughout the Northeast and beyond, the project team finds it necessary to provide a description of some important terms and concepts. The definitions below are drawn from the Forum's document *Glossary of Terms* (P. Horowitz 2009) unless otherwise noted.

Additionality: A criterion that says that avoided emissions should be recognized only for project activities or programs that would not have "happened anyway" in relation to a baseline estimate of project activity and associated with emissions reductions.

Adjusted gross savings: The change in energy consumption and/or demand that results directly from program-related actions taken by participants in an efficiency program, regardless of why they participated, after adjusting for such factors as installation and persistence rates and hours of use, among other possible factors. [Definition expanded from P. Horowitz (2009) on *gross savings*][26]

Baseline: Conditions, including energy consumption and related emissions, that would have occurred without implementation of the subject measure or project. Baseline conditions are sometimes referred to as "business-as-usual" conditions and are used to calculate program related efficiency or emissions savings. Baselines can be defined as either project-specific baselines or performance standard baselines (e.g. building codes).

Bias: The extent to which a measurement or a sampling or analytic method systematically underestimates or overestimates a value. Some examples of types of bias include engineering model bias; meter bias; sensor placement bias; inadequate or inappropriate estimate of what would have happened absent a program or measure installation; a sample that is unrepresentative of a population; and selection of other variables in an analysis that are too correlated with the savings variable (or each other) in explaining the dependent variable (such as consumption).

Construct Validity: "the degree to which an experimentally-determined definition matches the theoretical definition." (Jonas 2005) [See also *validity* below]

Deemed Savings: An estimate of energy or demand savings for a single unit of an installed energy efficiency measure that (a) has been developed from data sources and analytical methods that are widely considered acceptable for the measure and purpose, and (b) is applicable to the situation being evaluated. Individual parameters or calculation methods can also be deemed.

Free Rider: A program participant who would have implemented the program measure or practice in the absence of the program. Free riders can be 1) total, in which the participant's activity would have completely replicated the program measure; 2) partial, in which the participant's activity would have partially replicated the program measure; or 3) deferred, in which the participant's activity would have completely replicated the program measure, but at a future time than the program's timeframe.

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

Hours of use: The number of hours a day, on average, the device is in use; when gathered continuously through logging or metering use, such data may be used to build load shapes.

Installation rate: The number of measures incented by the program that actually get installed. [our definition]

[26] The exact nature of what might be included in adjusted gross is discussed in Section 2.2.1.

Leakage: Measure incented by the program installed outside of the program's jurisdiction. [our definition; this usage is distinct from that used by the air regulation community; see P. Horowitz 2009 for *leakage* in the context of air regulation]

Measure life: The life of an energy consuming measure, including its equipment life and measure persistence (not savings persistence). Measure life is not typically the absolute limit that a measure exists, but is instead commonly—but not always—measured by the median expected useful life of the measure. [Definition expanded from P. Horowitz (2009)]

Measure persistence: The duration of an energy consuming measure, taking into account business turnover, early retirement of installed equipment, and other reasons measures might be removed or discontinued.

Net savings: The total change in load that is attributable to an energy efficiency program. This change in load may include, implicitly or explicitly, the effects of [spillover], free riders, energy efficiency standards, changes in the level of energy service, and other causes of changes in energy consumption or demand. [Forum definition, but we substituted term "spillover" for "free driver" to be more consistent with terminology in this scoping paper.]

Net-to-Gross Ratio (NTGR): A factor representing net program savings divided by gross program savings that is applied to gross program impacts to convert them into net program load impacts. The factor itself may be made up of a variety of factors that create differences between gross and net savings, commonly including free riders and spillover. Other adjustments may include a correction factor to account for errors within the project tracking data, breakage, and other factors that may be estimated which relate the gross savings to the net effect of the program. Can be applied separately to either energy or demand savings.

Realization Rate: The term is used in several contexts in the development of reported program savings. The primary applications include the ratio of project tracking system savings data (e.g. initial estimates of project savings (the denominator of the ratio) to savings 1) adjusted for data errors, 2) that incorporate evaluated or verified results of the tracked savings (the numerators of the ratio). [The Forum definition also states that realization rate accounts for free ridership and/or spillover but we have removed this component as we found these components to be inconsistently used in realization rates. Moreover, the project team added slight clarifications to this definition.]

Rebound Effect: Also called Snap Back. A change in energy-using behavior that yields an increased level of service that is accompanied by an increase in energy use and occurs as a result of taking an energy efficiency action. The result of this effect is that the savings associated with the direct energy efficiency action is reduced by the resulting behavioral change.

Reliability: The quality of a measurement process that would produce similar results on: (1) repeated observations of the same condition or event; or (2) multiple observations of the same condition or event by different observers.

Spillover: Reductions in energy consumption and/or demand caused by the presence of an energy efficiency program, beyond the program-related gross savings of the participants and without financial or technical assistance from the program. There can be participant and/or nonparticipant spillover. Participant spillover is the additional energy savings that occur when a program participant independently installs energy efficiency measures or applies energy saving practices after having participated in the efficiency program as a result of the program's influence. Non-participant spillover refers to energy savings that occur when a program nonparticipant installs energy efficiency measures or applies energy savings practices as a result as a result of a program's influence. Additional distinctions may also be made to spillover such as "like" or "unlike" spillover (the former refers to the same measure as the program, while the latter refers to other actions taken as a result of program participation) and "inside" and "outside" of the project (the former referring to non-incented measures taken within the same project that had received an incentive while the latter is adoption of non-incented measures in an unrelated project). [Definition expanded from P. Horowitz (2009)]

Validity: "the extent to which [the measure] measures what it is intended to measure." (Rossi, Lipsey, and Freeman 2004:220)

2 Net Savings: Contextual Issues

This scoping paper focuses on the intersection of measurement approaches, policy needs, and consistency in the measurement of net savings. Before turning to these critical concerns, we find it necessary to describe the context surrounding the estimation and use of net savings. The contextual issues discussed here are not only essential background for the discussion that follows, but also served as key drivers for this scoping paper. The team addresses three contextual issues in the following section: 1) key audiences for net savings, 2) definitions of net savings, and 3) arguments for and against estimating net savings.

2.1 Meeting the Needs and Expectations of Diversifying Audiences for Net Savings Estimates

Historically, program administrators and energy regulators have been the main audiences for energy savings estimates, whether gross or net savings. These audiences typically used the information to assess how well programs were performing, to guide program revision and discontinuation, and to decide on rewards or penalties for program administrators (Saxonis 2007a, 2010; Messenger *et al*. 2010; In-depth interviews). The estimates, therefore, also influenced the activities of program planners and implementers, even if they were not among the primary audiences for the net savings estimates.

More recently, the audiences for energy savings estimates have grown to include new stakeholders and regulators with the expectation that energy efficiency efforts may result in substantial reductions in greenhouse gas emissions, while long-standing audiences have increased their scrutiny of estimates, largely in response to increased funding and expanded regional goals for energy savings and reduced emissions (Saxonis 2007a; In-depth interviews).

Table 2–1, in the left-hand column, lists the key audiences identified in the literature or by interviewees. The second column indicates whether the group is typically considered a primary audience (*i.e.*, the main audience for whom the estimates are developed) or a secondary audience (*i.e.*, an additional group that may use the estimate or be directly affected by it). The third column states whether the group is an existing or emerging audience, with existing audiences composed of those who have long been interested in and have used net savings estimates, and emerging audiences being those who have more recently become concerned about net savings. In two cases (policy makers and advocacy groups) the audience is listed as both existing and emerging because the literature and interviewees indicated that, while some members of these groups had always been interested in net savings (*i.e.*, a specific legislator or a low-income ratepayer advocate), attention to net savings within these audiences has recently increased. The final column states the major ways in which our sources and interviewees believe each audience uses net savings estimates. The primary message of the table is that net savings estimates are increasingly being developed with air regulators, legislators, and policy makers in mind, in large part because the net savings estimates are seen as critical components to making certain that

jurisdictions are meeting expanded goals for reductions in energy use and emissions. However, it must be noted that, at this time, it remains unclear whether air regulators will rely on adjusted gross savings or net savings to estimate emissions reductions.

While the existing and emerging audiences share some of the intended uses of energy savings estimates, their differing needs and perspectives contribute to differences in opinions regarding whether they require estimates of gross savings or net savings, and, if so, how to define and measure net savings, whether those definitions and methods should be consistent across the Northeast Region, the importance placed on net savings in assessing program performance, and even whether net savings should be measured at all. In short, the increased diversity of audiences is an underlying consideration in each of the topics discussed in the paper. Air regulators are one of the most critical new audiences for estimates of energy savings resulting from efficiency programs; therefore, we pay particular attention to questions raised by and concerns related to air regulators in this report.

Table 2–1: Audiences for Net Savings Estimates*

Audience	Primary or Secondary	Established or Emerging	Use of Net Savings Estimate
Energy efficiency program administrators and planners	Primary	Established	• Assess if program achieved savings goals • Identify strong and weak areas of program design. Redesign program based on understanding of strengths and weaknesses • Apply strong program designs for other products, in other jurisdictions • Recognize when to discontinue program
Energy efficiency regulators	Primary	Established	• Assess if program achieved savings or policy goals • Adjust payments to / funding of programs based on goal achievement • Recommend or require program revision or discontinuation based on goal achievement • Determine if the ratepayer / taxpayer funds are being spent cost effectively and wisely**
Air regulators***	Secondary, but increasingly primary	Emerging	• Disagreement over whether will require adjusted gross or net savings • Will apply emissions factors to energy savings to estimate greenhouse gas reductions • Assess degree to which efficiency programs have achieved greenhouse gas reduction targets
Legislators and other policy makers	Secondary, but increasingly primary	Established and Emerging	• Assess if program has achieved savings or policy goals**** • Determine if the ratepayer / taxpayer funds are being spent cost effectively and wisely*
Ratepayer / taxpayer advocacy groups	Secondary	Established and Emerging	• Determine if the ratepayer / taxpayer funds are being spent cost effectively and wisely*

* The ISOs and system planners are not included in this list as they require adjusted gross for the Forward Capacity Market. However, in the future, they may become an audience for net savings estimates.

** In the case of some programs—for example low income ones—cost effectiveness is less important than the wise or responsible expenditure of program funds.

***As discussed in Section 4.2, air regulators disagree over whether they will base estimates of emissions reductions on adjusted gross savings or net savings. Most argue that they will need energy savings estimated by time of day and time of year, perhaps for all 8,760 hours in the year.

**** Some legislators and policy makers interested in energy efficiency and renewable energy have been paying attention to program achievements, including net savings, for some time. However, recent increases in funding for energy efficiency programs and their ties to economic and job stimulus programs and to greenhouse gas reduction targets have expanded the number of legislators and policy makers paying attention to net program impacts. To offer an example, one state attorney general in the Northeast has given strong direction to program administrators on the types of data they can use to assess program impacts.

2.2 Definitions and Conceptualizations of Net Savings

On the surface, the interviewees and literature that we consulted from within the energy efficiency community reflect a common understanding of the definition of net energy savings. All cite some variation on the idea that they are the savings attributable to the program or savings that would not have occurred in the absence of the program. Scratching just below the surface, however, revealed substantial variation in the *operational* definition of net savings—that is, how one gets from gross to net savings. In this section we discuss the variation in definitions of net savings, and save the issue of consistent regional definitions for Section 5.

2.2.1 Net Savings and Free Ridership

Before turning to definitions of net savings, the project team finds it necessary to share one vital observation from our research. The Forum asked the project team to explore concerns about the conceptualization and measurement of net savings and about promoting consistency in net energy savings approaches throughout the Northeast Region. Therefore, we have used the terms "net energy savings" and "net savings" in this scoping paper. However, our investigation found that the literature—and interviewees—frequently focused their responses on free ridership measurement—specifically, the challenges surrounding its measurement. When addressing spillover, market effects, and other components that could be considered a part of net savings, the context was frequently in terms of criticizing a perceived regulatory preoccupation with free ridership at the expense of a broader view of net energy savings and program effects. This focus on free ridership underlies many of the concerns raised about the measurement of net savings and striving for regional consistency in approaches to net savings. Given the challenges of measuring free ridership and the fact that, by itself, it reduces the amount of energy savings attributed to a program, it is not surprising that the literature and interviewees tended to discuss its weaknesses and disadvantages, as opposed to its strengths and advantages.

The tendency to highlight the perceived shortcomings of free ridership follows directly from the energy efficiency community's current reliance on its quantification coupled with the limitations of methods used to estimate it (the latter of which is discussed in detail in Section 3). Regulators in many jurisdictions require the reporting of free ridership in an effort to ensure that ratepayer and taxpayer funds are being spent prudently. Yet for reasons discussed below in this scoping paper, the accurate measurement of free ridership—and spillover—remains challenging despite continual improvement in methods. If the energy efficiency community were to rely to a greater extent on approaches that do not focus on the separate quantification of free ridership, as opposed to net savings estimates that do not differentiate between free-ridership and spillover, it is likely that many of the concerns raised in this scoping paper would be less of an issue.[27] Such approaches include such as billing analysis, market-level approaches, historical tracing, and weight-of-evidence, among others.

[27] Although other concerns might become evident, as these estimation methods are not without their own disadvantages and challenges as discussed in Section 3.

2.2.2 Adjusted Gross Savings vs. Net Savings

The literature and interviewees named a number of energy savings parameters that could potentially be classified as components of net savings. These include not only free ridership, spillover, and NTGR, but also realization rates, measure persistence, installation rates, rebound effect, leakage, hours of use, and, for lifetime savings, measure life. We define these terms in Section 1.3.[28] Each of these parameters modifies the deemed or *ex ante* gross savings estimate based on the results of impact evaluations. However, over the course of our research, we came to differentiate these parameters into two groups. The first group includes what we view as likely components of *adjusted gross savings*: measure persistence, installation rates, leakage, hours of use, measure life, and realization rates (which may include other components). The second group of parameters includes what we argue are the appropriate components of *net savings*: free ridership, spillover, NTGR, and the rebound effect. Ideally, program administrators would first estimate adjusted gross savings, and then apply free ridership and spillover rates or NTGR to adjusted gross savings to estimate net savings. Therefore, the final net savings estimate does take the components of adjusted gross savings into account.

The distinction between the parameters of adjusted gross savings and net savings rests on two factors: 1) the literature and interviewees focused their attention on free ridership, spillover, and NTGR, and 2) the parameters we tie to adjusted gross savings are at least theoretically amenable to direct measurement or calculations derived from direct measurement through metering use or observation of the measure or behavior. In contrast, the net savings parameters cannot be directly measured because they are at least partially based on a counterfactual—that is, on "what would have happened," not what actually did happen. This is the reason that net savings estimates can be controversial, as we discuss in Section 3. It is important to note that this distinction has some precedence in the Northeast: currently program administrators must report only adjusted gross savings to the Independent System (ISOs), although the exact parameters used in their estimates of adjusted gross savings may be somewhat different than those we discuss here.

2.2.3 Inputs into Net Savings Calculations

The first critical area of difference in how net savings are estimated involves the inputs used to calculate net energy savings. As mentioned above, there are two generally recognized—though by no means universally measured—components of net savings: free ridership and spillover. The estimation of these two components provides one approach for deriving the NTGR, although this ratio can also be developed at the market level, which captures the effects of free ridership and spillover but does not allow for their separate calculation. Nearly all programs or jurisdictions that require the calculation of net savings adjust for free ridership, but only some also adjust for spillover. For example, as shown in Table 2–2, Messenger and his colleagues (2010) found that ten of the 15 jurisdictions included in their study calculated net savings; all ten adjusted for free

[28] Based on our interviews, neither leakage, rebound, nor spillover are regularly included in estimates of adjusted gross or net savings in the Northeast, although other jurisdictions such as California measure leakage and British Columbia measure rebound.

ridership, but only eight adjusted for spillover. Of these eight, three adjusted for spillover in just "a few or some cases." Although Messenger and his colleagues did not survey individuals from all Forum states, Table 2–2 makes clear that the Northeast Region states that were included differ in their inclusion of free ridership and spillover. Specifically, Pennsylvania makes neither adjustment, Maine adjusts for free ridership but not spillover, Connecticut also adjusts for free ridership but spillover only sometimes, and Massachusetts and New York adjust for both.[29]

Table 2–2: Market Influences and Program Effects Included in Estimates of Net Savings from Energy Efficiency Programs*

State	Free-ridership	Spillover/ Market Effects
CA	Yes	Yes in few cases
CT	Yes	Yes in some cases
FL	Yes	Yes
IA	No	No
ID	No	No
IL	Yes	Yes
MA	Yes	Yes
ME	Yes	No
MN	Yes	No
NEEA	No	No
NY	Yes	Yes
OR	Yes	Yes
PA	No	NA
TX	No	No
WI	Yes	Yes in few cases
Total Yes	***10***	***8***

* Source Messenger *et al*. 2010. The original table also included a column on leakage to other states, but only one state—California—counted such leakage and for just selected programs.

To complicate matters further, jurisdictions and programs sometimes distinguish among different types of free ridership or spillover.[30] While most accept the idea of a full free rider—a participant who would have adopted the exact same measure or behavior, at the same time, and for full price (if applicable) in the absence of the program—opinions differ regarding a partial free rider—a participant who would have taken some action to bring about energy savings, but perhaps at a somewhat reduced level of efficiency or at a later period in time (this last is also known as a deferred free rider).[31] Not all jurisdictions recognize partial free ridership, and, even when they

[29] While Massachusetts adjusts for both, one interviewee explained that the state accepts market-level estimates of NTG ratios for market transformation programs instead of requiring separate, isolated estimates of free ridership and spillover. Our experience performing evaluations in Connecticut and New York confirms that they also do not always require the separate reporting of free ridership and spillover for market transformation programs.

[30] Definitions in this section come from multiple sources, most notably Saxonis 2007a, Titus and Michals 2008, and Skumatz *et al*. 2009. However, interviewees and other literature sources also provided similar definitions. We developed the examples based on our experiences in energy efficiency evaluation.

[31] For example, an appliance incentive program may have induced a household to buy a Tier 3 clothes washer instead of a Tier 2 model, and to make that purchase a few months earlier than originally planned.

do, the concept may be defined differently across programs or jurisdictions, often for very good reasons. For example, it may make sense to limit deferred free ridership for a residential central air conditioning program to purchases that would have been made in the past year, but the period considered for deferred free ridership for a large HVAC retrofit in a commercial building may justifiably be longer, in acknowledgement of the longer planning and budgeting cycles in the business world.

Likewise, jurisdictions differ not only in their inclusion of free ridership, but also in whether or not they allow non-participant and participant spillover in the calculation of net savings, with participant spillover sometimes being further categorized as inside the project and outside of the project. Non-participant spillover occurs when someone takes an energy-saving action because of a program but did not directly take part in the program. For example, a program may influence a builder to decide to adhere to the specifications of an ENERGY STAR Home but not to take advantage of any training, certification, or incentive offers associated with the program. Participant spillover inside a project includes additional actions taken at the same site, while participant spillover outside a project occurs at a different site (Megdal, Patil, Gregoire, Meissner, and Parlin 2009). A small business owner who takes part in a commercial lighting program may be influenced to install ENERGY STAR-qualified office equipment in the original project site (inside project spillover) and both types of products in another site (outside project spillover), all without further engaging the program or any others offered by the same program administrator.

Furthermore, while some jurisdictions require that net impact assessments include free ridership, participant spillover, and/or non-participant spillover, not all require that these components be calculated separately. Massachusetts, for example, includes all types of free ridership and spillover in its definition of net savings but does not *require* that these components be reported separately, especially for programs with a market transformation focus. The state's approach reflects a recognition that it is extremely challenging to isolate free ridership and spillover in many market transformation-type programs, particularly those that rely on manufacturer and retailer markdowns and buydowns, a topic to which we return in the next section.

2.2.4 Net Savings Equations

In addition to the differences in which adjustments they make to gross savings estimates, programs and jurisdictions also turn to divergent calculations of net savings. The most common calculation of net savings (*e.g.,* Rathbun, Sabo, and Zent 2003; Skumatz *et al*. 2009; NMR Group 2010), is denoted below by the following net savings equation:

$$\text{Net Savings}_1 = \text{Gross} * (1 - \text{Free Ridership})$$

or its spillover-inclusive alternative (assuming a combined participant and non-participant spillover estimate):

$$\text{Net Savings}_2 = \text{Gross} * (1 - \text{Free Ridership} + \text{Spillover})$$

The portions of these equations in parentheses comprise the NTGR. The "gross" could be adjusted gross, and, if so, the equations would already take into account such parameters as installation rates and hours of use. If, however, the "gross" is the deemed or ex-ante gross savings estimate, one would also need to apply these other parameters—or an overall realization rate encompassing them—to yield both adjusted gross savings and net savings. Rathbun and her colleagues (2003), for example, suggest Net Savings = Gross * (Realization Rate) * (1 – Free Ridership + Spillover).

Skumatz and her colleagues (2009) also suggest that the energy efficiency community explore the relative strengths and weaknesses of using the equation above versus the following equation, which some commentators argue was once used more frequently than it is now:

$$\text{Net Savings}_3 = \text{Gross} * (1 - \text{Free Ridership}) * (1 + \text{Spillover})$$

Table 2–3 compares the net savings estimates resulting from these three equations, using the same assumptions about gross savings, free ridership, and spillover. Of course, the results would diverge more if we varied assumptions about partial free ridership and types of spillover. It is worthwhile noting that Keating (2009) has shown that multiplicative approaches such as in the third net savings equation and those in which one multiplies various components of free ridership together to arrive at an estimate of free ridership or net savings create a downward bias on the estimate. This results from the mathematical properties of multiplying decimal places and should only be used in limited situations (such as when the designation as free rider is conditional on other factors being true or the method to estimate spillover does not consider attribution). To illustrate his point using the assumptions in Table 2–3, the result of (1 – Free Ridership + Spillover) or (1 – 20% + 30%) equals 1.1 while that of (1 – Free Ridership) * (1 + Spillover) or (1 – 20%) * (1 + 30%) = 1.04, demonstrating the downward push of the latter method. Changing the free ridership or spillover assumptions to other values yields the same impact—a consistently lower estimate with the multiplicative method.

Table 2–3: Net Savings Estimate Using Three Equations

Assumptions: Gross = 150 units*, Free ridership = 20%, Spillover = 30%	
Net Savings Equation	**Net Savings**
NTG_1 = Gross * (1-Free Ridership)	120 units
NTG_2 = Gross * (1-Free Ridership + Spillover)	165 units
NTG_3 = (1-Free Ridership) * (1+Spillover)	156 units

* Units could be kWh, therms, or any other common measure of energy consumption

The characteristics of some program designs and products, as well as data availability, exacerbate the often substantial challenges of measuring free ridership and spillover. This is particularly the case with upstream programs, of which CFL programs are the quintessential, but by no means the sole example. In response to these challenges, program administrators and evaluators have turned to market-based approaches to estimating net savings, even though these approaches do not always allow for separate estimation of free ridership and spillover (*e.g.* Goldberg *et al*. 2006; Winch and Talerico 2008; NMR Group 2010). In their most basic form,

market-based approaches rely on the following equation (with the NTGR being the portion inside the brackets):

$$\text{Net Sales}_4 = \text{Program-supported sales} * [(\text{Total Sales} - \text{Baseline Sales}) \div \text{Program-supported sales}]$$[32]

While this approach works well when applied to market transformation programs and when consistent sales or purchase data exist, it can be controversial. First, it usually cannot provide separate estimates of free ridership and spillover, limiting its acceptance in jurisdictions accustomed to seeing such estimates or actually requiring them. Second, as with nearly all methods of measuring net savings, determining the inputs for total and baseline sales—and sometimes for program-supported sales—can be challenging and controversial, a topic we address in Section 3 on methodological issues below.

2.3 Arguments for and against Measuring Net Savings

The energy efficiency literature and the in-depth interviews provide a number of arguments for and against measuring net savings. The arguments in support of measuring net savings include the following:

- Establishes that the program brought about energy savings that would not have occurred in its absence
- Quantifies the amount of savings achieved as a result of the program in a specific time period
- Assesses the effectiveness of various program designs and whether the designs should be replicated, expanded, revised, or discontinued
- Ensures that ratepayer or taxpayer funds are being spent responsibly and in a manner that ensures that efficiency truly is the lowest-cost resource and not just claimed to be
- Uncovers fraudulent program implementation practices such as the claiming of savings from installations that occurred prior to any interaction with the program
- Complies with regulatory requirements to report estimates of net savings along with gross savings

[32] In place of sales, researchers may also insert household-level purchases/incented products or base the equation on energy savings instead of sales or purchases.

The arguments presented against measuring net savings can be summarized as follows:

- Focuses too heavily on narrowly defined metrics of individual program success or failure—especially free ridership—while deemphasizing other important impacts, such as non-energy benefits and behavioral effects, as well as portfolio and policy-level impacts
- Creates the impression that the estimates accurately represent the savings attributable to the program when, in reality, the methods available often yield results that may not accurately represent actual program-induced savings due to shortcomings related to bias, reliability, and validity
- Requires large expenditures of resources that are not in keeping with the importance of the estimates and their reliability and validity, thereby diverting resources from other planning, evaluation, and implementation activities that could yield greater benefits

We expand on these arguments for and against the measurement of net savings below.

2.3.1 Arguments for Measuring Net Savings

We have identified six primary arguments for measuring net savings.[33] The first and most common argument is that it establishes whether the program *caused* energy savings or demand reduction that would not have otherwise occurred in the absence of the program. Measuring net impacts—and their individual components of free ridership, participant spillover, and non-participant spillover—allows the program to state with at least some certainty that it was responsible for the creation of these energy savings above and beyond what would otherwise have occurred.

Similarly, the second argument asserts that measuring net savings not only demonstrates that the program caused the savings but also quantifies the amount of savings actually due to the program. Although some commentators questioned the validity of estimates of net savings, they agreed that this was one of the reasons that programs measure net savings.

The third argument for measuring net savings is that the results help identify the program designs that are most effective, as well as the programs that are not performing well. This allows program administrators to know which programs to keep as is, which programs may require revisions (*e.g.* in incentive levels, qualifying equipment and efficiency levels, targeted customers, marketing, product mix, *etc*.) to limit free ridership and increase savings, and which programs or measures should be dropped from the portfolio altogether. As one energy regulator expressed, “If there is a large percentage of free ridership, then the program should be redesigned or replaced. It’s not so much knowing the savings but how big the free ridership is.” In a related corollary, some commentators noted that the effectiveness of program designs can change over time, such that programs require relatively frequent monitoring so they can take credit for additional savings when free ridership is low and spillover high, but swiftly respond to market

[33] Sources for “reasons to measure” come primarily from in-depth interviews; TecMarket Works 2004; Rossi *et al*. 2004; Howell and Yemane 2006; Saxonis 2007a; Hoefgen, Li, Azulay, Prahl, and Oman 2008; Skumatz 2009; Messenger 2010, and Drew 2010. Numerous other sources also touch on these issues.

changes in order to protect savings when free ridership increases and spillover decreases, which might signify that market is nearly transformed. It is also the case that net impact estimates are often used to adjust deemed savings used for program planning and interim reporting, ideally bringing the deemed and achieved savings closer together.

The fourth argument in support of net savings measurement is that it ensures that ratepayer or taxpayer funds are being spent responsibly and in a manner that helps bring about a reliable energy supply. Most energy efficiency programs are funded through fees attached to utility bills or via government programs (or some combination of the two). It is the responsibility of program administrators to make sure that they are spending these funds in a manner that benefits the public by achieving savings that would not have otherwise occurred without the expenditure of these funds. If net savings are low, then the program administrators should take actions to ensure that the ratepayer and taxpayer monies are put to more effective uses. The Electric Consumers Resource Council (ELCON, 2008) suggests a corollary to this perspective. Energy efficiency, ELCON argues, is often presented as the lowest cost "first fuel." Estimating net savings—and cost effectiveness—can help determine if it truly is the lowest cost alternative *from a consumer's perspective*, not the perspective of the suppliers or generators. They worry that without proper accounting of net savings, consumers will pay twice—once for energy efficiency measures and, if the measures fall short of expectations, again for building new generation sources to provide the supply that energy efficiency promised but did not deliver.

The fifth argument for measuring net savings is to uncover fraudulent program implementation practices. Cases have been documented in which program implementers have claimed savings from activities that they clearly did not influence. These include claiming the savings from installations that occurred prior to any interaction with the program or decreased usage from changes in activity such as dropping a shift in a factory or having someone move out of the house.

Finally, the sixth commonly cited reason for measuring net savings is that it complies with regulatory requirements to report net savings along with gross savings. Regulators explain that they use net savings to make sure that the public good is being met, energy supply is stable and reliable, and greenhouse gases and other emissions are being reduced. In some jurisdictions (*e.g.* Massachusetts and California), policies mandate that regulators reward or penalize programs financially based on their net savings achievements.

2.3.2 Arguments against Measuring Net Savings[34]

Although net savings are commonly measured, some commentators argue that they should not be measured at all, that they should only be measured in limited situations, or that they should not be applied in the ways that they are currently. If fact, Messenger and his colleagues (2010) found

[34] We remind the reader that these arguments tend to be more about the individual components of net savings—free ridership and spillover—than about the broader concept of net savings. As discussed in Section 5.4 many of these arguments would become moot if most current approaches did not rely on the quantification of free ridership, which is a contentious practice.

that only 63% of the 50 respondents to their survey (made up of state regulators, policy makers, program administrators, and local evaluation practitioners from across the United States) believed that net savings should be reported, either together with gross savings (57%) or alone (6%). Of the remainder, 15% supported reporting gross savings alone, but 22% either said "don't know" or "neither should be reported." The commonly cited reasons against measuring net savings generally fall into the three categories, discussed below. It should be noted, however, that while a few sources take a hard line against measuring net savings based on its limited ability to take into account broader accomplishments, most of the commentators who voice reservations about measuring net savings argue instead that net savings should be de-emphasized as the primary factor in determining program success or failure and in calculating the concomitant payments to programs that come from this determination.

The first category of reasons against measuring net savings rests on the premise that the goals of energy efficiency programs extend beyond short-term energy savings; they include market transformation and, increasingly, other non-energy benefits as well (In-depth interviews; Eto, Prahl, and Schlegel 1996; Titus and Michals 2008; Friedmann 2007).[35] Holders of this opinion argued that the regulatory emphasis on net savings—particularly free ridership—fails to capture the extent to which programs have broader market effects (*e.g.,* increasing the number and diversity of efficient products on the market), bring about non-energy benefits (*e.g.,* reducing greenhouse gas emissions, increasing social justice), or affect attitudinal and behavioral changes that potentially will influence future energy-related decisions. A program administrator provided a vivid analogy of the narrow focus on net savings at the expense of broader program accomplishments:

> *We are looking at the freckle on the back of the elephant. We measure [net savings] program by program and measure by measure, yet there are so many programs that this way of looking at it doesn't tell us what we are doing at a high level. This is a legacy from shareholder incentives and public policy perspectives on programs. We should move away from these policies and look at a higher level as to what we are accomplishing…. [Measuring net savings] diverts our attention from other things.*

In a related argument, one interviewee explained, "It's dangerous to take [net savings] too seriously because there are so many factors that affect net savings that aren't in the program planner's control." In other words, high free ridership or low NTG ratios may be the result of factors that the program cannot address, such as the stage in the development of the market at the time of the program or competing influences in the marketplace. High free ridership may even indicate that the program is behaving as intended. As Friedmann (2007) writes, "Free ridership might not be a bad thing in a market transformation world." Efficiency programs exist to increase the market share of efficient equipment (or the uptake of efficiency behaviors). It is very likely that, as market share increases, people who are not directly influenced by the program

[35] Others would counter that energy savings is the goal of market transformation, and net savings is one of the primary ways of assessing if market transformation has occurred (*e.g.* if free ridership is high).

adopt the measure in response to the general market conditions created by the program. The very conditions that the program has sought to create—and succeeded in creating—may count against it if assessments of the program's effectiveness focus too narrowly on free ridership and do not adequately account for spillover or the cumulative effects of prior program activity on current measure adoption. In such situations, to discontinue the program (or to penalize the program administrators financially) would be more detrimental than allowing it to continue.[36]

The second category of reasons for not measuring net savings is methodological (In-depth interviews; Saxonis 1995; Titus and Michals 2008; Skumatz *et al.* 2009). Because estimating net savings requires "measurement" of a counterfactual—what would have happened in the absence of the program (see Section 3)—all methods involve a degree of uncertainty and bias, albeit some more than others (Rossi *et al.* 2004).[37] The methods most frequently available to energy efficiency program evaluators—participants' responses to survey questions, reliance on a non-randomized control group, *etc.*—are often considered unreliable, raising concerns about the accuracy of the resulting net savings estimates. Further, most of the individuals we interviewed asserted that, as spillover is more difficult to capture than free ridership and thus is either underestimated or not measured, and as free ridership is often overestimated, net savings is frequently undercounted. Inaccurate results can lead to false conclusions, causing regulators and administrators to respond in a manner that is contradictory to the true, but still unmeasured, net savings of the program, allowing weak programs to remain active and forcing strong programs to cease (Saxonis 1995).

Self reporting of free ridership and spillover through participants' responses to survey questions has its own set of methodological issues related to timing. Typically, free ridership and spillover are estimated at the same time, usually many months after the individual has taken part in the program. Often, respondents no longer remember what drove them to participate in the program; sometimes they have reframed their decision psychologically, convincing themselves they would have taken the step without the program (In-depth interviews; Peters and McRae 2008). This leads to overestimates of free ridership. Saxonis (2007b) also notes that timing issues are important for spillover so that adequate time can pass for end users to take additional steps, but not so much time that linking the action to program participation becomes tenuous.[38]

---

[36] To offer one example, a mounting body of evidence points to decreasing NTG ratios for CFL programs across the nation, in part because the results are based on comparisons of program areas to non-program areas with very little differences in CFL sales. Some commentators believe that this points to the need to revise or discontinue CFL programs, but others believe that long-standing program states, such as those in much of New England, have become "addicted" to the incented program bulb price and would not buy them at the unsupported prices offered in non-program areas. Evaluations currently being conducted in areas that have revised their CFL programs may shed some light on this debate.

[37] As discussed in Section 3, methods relying on random assignment or on objective data (*e.g.* sales data), are *usually* considered more reliable than quasi-experimental methods (*i.e.* the use of a non-random comparison group) or those relying on participant self-reports of actual or counterfactual behavior.

[38] See Section 3.1 for an in-depth discussion of concerns related to self-reported approaches, including more on the psychological factors that affect how respondents answer questions about the influence of the program on their actions and behavior.

The final set of arguments against measuring net savings is that doing so requires a great deal of time and money, and the level of effort is often not commensurate with the quality or significance of the results (Titus and Michals 2008; Peters and McRae 2008). In other words, a lot of resources may be spent on gathering data that are of only marginal use to the program administrators and evaluators. Commentators also noted that, in some cases, the program savings are so large—or free ridership so low—that net savings will remain high, and that spending resources on measuring free ridership in these circumstances may not be justified (Saxonis 1991a).

2.3.3 Net Savings-Based Payments to Program Administrators

As mentioned in Section 2.3.1, some jurisdictions reward or penalize programs based on net savings. This practice is one of the more contentious issues addressed in the literature and in-depth interviews because the amount of money in question can sometimes be quite large even while the results involve a good deal of uncertainty. Importantly, while there was a tendency for regulators to support this practice and program administrators to oppose it, at least one net-savings expert who is neither a regulator nor a program administrator argued against penalizing programs for not meeting goals based on estimates of net savings. He likened the practice to "shooting your allies…. If you're trying to sell the product you work with [your allies] to push the market forward. You don't punish [them]." The same interviewee also asserted that regulators needed to accept free ridership as a form of "overhead cost" and de-emphasize its importance in assessing program success or failure. This individual suggested that targets and rewards be set in terms of gross savings, not net, even though net may still be useful for purposes of program planning and revision.

Other commentators, largely regulators, disagreed.[39] They argued that it is the responsibility of regulators to ensure that ratepayer and taxpayer funds are being used wisely. Paying for practices that would have occurred without the use of these funds is not a wise use of ratepayer and taxpayer monies. Penalizing poor performance stimulates program planners to rethink their designs and come up with more effective ways to use ratepayer and taxpayer funds, just as rewarding strong performance leads to replication of approaches that provide the greatest "bang for the buck."

Ridge, Willems, Fagan, and Randazzo (2009) question not so much the practice of rewarding or penalizing program administrators based on net savings but instead the negative reaction that some policy makers and regulators have to the uncertainty inherent in estimating net savings:

> *What policy makers and regulators sometimes forget is that no measurement system, no matter how rigorous, within the broader evaluation community, can meet [many of their desired levels of precision]. When evaluators fail to deliver the level of accuracy and*

[39] There is a tendency for program administrators and net savings experts to hold the opinion that payments to programs should not be based on net savings. Not surprisingly many—but by no means all—energy and air regulators are of the opinion that such payments *should* be based on net savings.

> *precision required by regulators, one should not flog the evaluators and condemn their evaluation methods. Rather, one should change the regulatory framework from the high-stakes system of rewards and penalties so that they are more consistent with best evaluation practices (2009:142).*

Energy efficiency is not alone in struggling to provide impact evaluation and net effects results that meet the levels of certainty, reliability, and validity desired by regulators, funders, or other overseers of various social, health, and environmental programs. Other fields of evaluation face similar issues, a topic to which we now turn.

2.4 Measuring Net Impacts in Other Fields

The project team performed a limited investigation into the question of how other fields of program evaluation address concerns about net impacts and how often they are expected to measure net impacts, if at all. While other fields do not have to measure net *savings*, they are often expected to show that their program has produced intended outcomes (*e.g.*, increased standardized test scores, reduced rates of HIV infection, preserving the number of spotted owl nesting pairs, *etc*.) that would not have occurred in the absence of the program. Sometimes this documentation takes the form of measuring net impacts, but other times the efforts simply measure program performance and gross outcomes. While documenting the methods of estimating net impacts and the frequency with which they are measured in other areas of evaluation in detail was beyond the scope of the project, we conducted a limited literature review on these topics, made inquiries to a few colleagues in other areas of evaluation, and drew on our own experiences in evaluation, monitoring, and verification in other fields.

2.4.1 Net Impacts Measurement in Other Areas of Program Evaluation

Perhaps the most important finding from our limited exploration of net impacts in other fields is that energy efficiency is far from alone in its struggle to measure program impacts. Programs often have to demonstrate that their efforts have led to the intended outcomes. Furthermore, the literature on how to perform program evaluation supports the practice of documenting net impacts; text books, journal articles, book chapters, and meetings papers are replete with discussions and debates on how best to measure net impacts (*e.g.*, Fitz-Gibbon and Morris 1987; Mohr 1995; Connell and Kubisch 1998; Mayne 2001; Rossi *et al.* 2004; Howell and Yemane 2006; Cook, Scriven, Coryn, and Evergreen 2010). As Rossi and his colleagues (2004) explain, "The challenge for evaluators…is to assess not only the outcomes…but also the degree to which any change in outcomes is attributable to the program itself."

While measuring net impacts may be seen as "best practice" in the broader evaluation community, in reality it seems that program administrators and evaluators differ in the degree to which they actually measure net impacts for their funders or clients. For example, Ferraro (2009) explains that many administrators and evaluators of environmental programs state that, unlike social policy programs, they should not or cannot measure what would have happened in the

absence of the program because of the complexities of doing so in the natural world. Instead, they should only monitor what happens after the intervention. Ferraro rejects this argument and calls for increased use of experimental and quasi-experimental designs for assessing the impacts of environmental programs, but the point remains that, currently, environmental programs rarely estimate net impacts. Turning to another field, in response to our direct inquiries, an evaluator in the health services community as well as an individual who develops evaluations for a national social marketing firm reported that most of their colleagues and clients focus more on gross impacts and documenting outputs (*e.g.,* counting patients served) and less on net impacts. The social marketing evaluator qualified his answer by saying that the Centers for Disease Control (CDC) and the National Institute of Health (NIH) measure net impacts "because they promote evidence-based interventions." These two individuals, then, agreed that not many of their clients desire net impacts, but only one of the two pointed to frequent usage of the approach in at least some sectors of the health services community.

It is also the case that academic researchers conduct program evaluations in which they seek to determine the net impacts of programs or policies, sometimes as independent researchers, other times as contractors working for program administrators or funding agencies. For example, a team lead by researchers at the University of Minnesota, partnering with the Minnesota Department of Public Health, evaluated the impact of Medicare coverage policies on provider behavior and health care utilization (Foote, Virnig, Town, and Hartman 2008). The authors found that the policies did not systematically affect either provider behavior or health care utilization—that is, they could identify no net program impacts on these two hypothesized outcomes. Neugart (2008) relied on an "agent-based computational economic (ACE) model" to evaluate the impact of government-financed training measures on moving individuals from unemployment to employment. His research found that such programs do have the intended effect of moving participants into jobs, but this comes at the expense of those who do not participate, because they get jobs at lower rates than they would have had the program not existed. Drawing on our own experience, the evaluators working on this paper have themselves been involved in evaluations of community development, education, public health, and environmental programs in which we or our colleagues have had to document the net impacts of programs, but we also have worked with programs in which we simply monitored the immediate outputs and gross impacts of the programs and did not seek to measure net impacts.

In an effort to avoid the frustrating conclusion that "it depends," we offer the following generalization: it seems as if the size of the program and its budget are positively correlated with the likelihood that it will undergo full net impact evaluation, but the frequency of net impacts evaluation depends on the nature of the program and how politically sensitive the project is. Large public health, social services, and education programs funded by the federal government appear to be among those most often expected to conduct net impacts evaluation, but again we stress that this is based on a limited exploration of practices in other fields.

2.4.2 Timing of Net Impacts Assessment in Other Areas of Evaluation

The project team was also asked to explore the question of the frequency of measuring net impacts. The program evaluation literature draws a distinction between performance measurement and impact evaluation. Performance measurement focuses on tracking key program outputs or indicators such as number of meals served, teachers trained, or children immunized. The data are constantly tracked and periodically reviewed to make sure the program is functioning as intended. In fact, since the passage of the Government Performance Results Act of 1993 (GPRA), all federally funded programs have had to engage in annual performance measurement. The effort was expanded with the 2006 launch of "Expect More" and its Program Assessment Rating Tool (PART).[40] While the Obama Administration has updated PART, the movement is toward greater accountability and reporting, not less (DeGroff, Schooley, Chapel, and Poister 2010).

While some sources we reviewed advocated expanding performance measurement efforts to address issues of attribution,[41] particularly on a qualitative basis and tied closely to program theory (*e.g.* Connell and Kubisch 1998; Mayne 2001; Blumstein 2009; DeGroff *et al*. 2010), the call is not universal. Instead, more program evaluators argued that performance measurement—without attention to attribution or net impacts—should happen on a fairly constant basis, while impact evaluation should be conducted at longer intervals, with the timing dependent on the nature of the program and its duration. Often, impact evaluation occurs toward the end or a program's life, or even after it has ended. This does not mean that evaluators are not working with the program while it is operating—and even during the planning phases. In fact, in the ideal situation, process and impact evaluators are present at the program planning stages and work with the program administrators to make sure program design and data collection are performed in a way that is conducive to evaluation. More typically, process evaluation occurs while the program is operating, and impact evaluation at the end of the program period.

We found two noteworthy exceptions to this tendency, both mentioned by Cook during a debate with Scriven (Cook *et al.* 2010). We admit that one could quibble that these examples are not pure, but they are illustrative. The first example involves programs stemming from the federal No Child Left Behind (NCLB) Act and similar state initiatives.[42] Every year, states measure the performance of children on standardized tests. Net impacts are assessed by comparing the test scores of the school to itself over time and to the scores from other schools over time and in that particular year. Of course, one problem with calling this approach "net" is that *every school in the state receiving public funds takes part in the program*, so there is no comparable control group. However, it is also the case that many evaluators at the district, state, and federal levels as

[40] See the Expect More website at http://www.whitehouse.gov/omb/expectmore/index.html and links to various Government Accountability Office and Office of Management and Budget publications at the Centers for Disease Control Website http://www.cdc.gov/eval/resources.htm#docs_gpra.

[41] That is, assessing not only the net impact of the program but also separating its impacts from those of other programs, a situation increasingly common in energy efficiency as we address in Section 4.3

[42] Ample literature exists documenting the strengths and weaknesses of NCLB on the education of children. Here we focus purely on the measurement issues, and do not address the controversies surrounding the program.

well as third party evaluators conduct assessments in which they attempt to isolate the effects of the program from other factors, often using statistical models. The second example involves the testing of new medications and procedures in the health-care industry—what are often called medical trials. While it may be a stretch to classify trials as "program evaluation," the important point is that such testing includes constant performance measurement and typically involves a comparison or control group—often a randomly sampled one.[43] Because patient health is at risk, the results are monitored carefully; if early results point to substantial problems or exceptionally positive results, the trial may be stopped and the results acted upon immediately.

Based on our limited research, it seems as if energy efficiency is one of the few fields requiring frequent assessments of net impacts. In other fields, impact assessment often occurs at longer intervals of time, sometimes not until the program has ended.[44] In place of frequent impact evaluations, other fields use performance measurement to provide the quick feedback needed to know whether to stay the course or to revise the program. But there is a caveat: programs expected to have immediate effects, that are controversial, and/or that are sensitive to outside influences are often subject to more frequent impact evaluations—or at least more in-depth performance measurement and comparison to control groups. In fact, we see the same practice in energy efficiency: upstream CFL programs are often subject to annual or biannual impact assessments, but commercial new construction programs sometimes undergo impact evaluation at less frequent intervals.

[43] The use of the control group for medical trials can lead to ethical concerns, and for this reason the research is often overseen by human subject review panels. These panels sometimes allow quite invasive procedures—sometime even brain surgery—on the control group in order to assess the net impact of the practice.

[44] Because many environmental, social, public health, and educational programs are funded by grants or are time-sensitive in nature, they frequently have defined start and end dates. In contrast, while energy efficiency programs typically have funding cycles, they usually are expected to continue indefinitely until the program is no longer effective or the program administrators, regulators, and sometimes legislators decide to spend the money in a different manner.

3 Advantages and Disadvantages of Current Net Savings Estimation Methods

The challenges of measuring net energy savings are not new. Many of the conceptual and methodological critiques raised today echo those of twenty years ago, despite the fact that methods of measurements have greatly improved in the interim (TecMarket Works 2004; Saxonis 2007a). Although some methods of calculating net savings provide more statistically reliable and valid estimates than others, the attempt to measure net savings will always embody some degree of uncertainty and bias, precisely because it forces evaluators to estimate the "counterfactual," something that never happened. As we mentioned in Section 1, this paper does not present a detailed discussion of the advantages and disadvantages of approaches to estimating net savings, as other authors have previously provided such assessments.[45]

Table 3–1, which begins on the next page, summarizes the key approaches and their advantages and disadvantages and provides an example of a study or situation in which the approach has been used. It should be kept in mind that not all advantages and disadvantages are equally important; for example, Table 3–1 lists only one advantage and multiple disadvantages for econometric modeling (*i.e.*, billing analysis), but that single advantage—that the approach can provide statistically valid estimates of net impacts in many situations—is a very important one that sometimes outweighs the disadvantages. On page 33, we provide a brief discussion of some of the most critical concerns raised about common approaches to measuring net savings.[46]

[45] For comprehensive reviews see Skumatz *et al.* 2009; Rosenberg and Hoefgen 2009; and Hoefgen 2010. Other authors discuss the advantages and disadvantages of individual approaches, such as Goldberg *et al.* 2006 and Saxonis 1991a, among others.

[46] It should be noted that many studies actually rely on hybrid approaches, drawing on one or more of the methods described in this section and Table 3–1. To offer one example, the recent multistate CFL modeling effort relied on self-reported estimates of purchases rather than sales data but ultimately turned to a modeled cross-sectional approach (NMR 2010).

Table 3–1: Net Savings Estimation Methods

Method	Definition	Advantages	Disadvantages	Example*
Randomized, controlled experiments	Random assignments of individuals to a participant (treatment) group or non-participant (control) group; methods used to estimate NTG could include econometric modeling, cross-sectional approaches, and others not included in this table; rarely used in energy efficiency	• Limits bias • Increases reliability and validity • Widely accepted in natural and social sciences as the "gold standard" of research designs	• Bias can result if random assignment occurs among volunteers or if "drop out" rate in study differs on key characteristics • Ethical concerns about assigning some ratepayers / taxpayers to control group • Cannot be applied after the program is implemented, and must be incorporated into program planning	Allcott and Mullainathan (2010) on OPOWER programs implemented in multiple states throughout the nation; applied billing analysis and simpler comparison methods to assess effects of this randomized, controlled experiment
Deemed or Stipulated Estimates	A specific NTG ratio is assumed and applied to the program(s); generally negotiated between program administrators and regulators or their designates often based on results of prior NTG studies	• Simplest and least expensive	• Potential for incorrect estimate • Not based on information that is specific to the program or program period	CPUC uses deemed savings (listed in its DEER database) for planning purposes and interim savings estimates for its programs; updates deemed savings based on results of NTG studies
Historical Tracing, *Modus Operandi*, and similar approaches	Uses information from a wide range of sources; reconstructs events leading to an outcome of interest to develop a 'weight of evidence' conclusion regarding influence of program	• Draws upon multiple information sources	• Difficult or impossible to determine magnitude of effects • Cannot assign statistical precision	Rufo (1999) on California's Nonresidential Standard Performance Contract Program

Method	Definition	Advantages	Disadvantages	Example*
Econometric Modeling – typically billing analysis (regression analysis of billing data)	Models participant and non-participant historical billing data to calculate annual net energy and demand savings; models compare participants' and non-participants' energy and demand consumption, controlling for external variables that could influence changes in use	• Can provide statistically valid estimates of net impacts[47]	• Data not always available for time period needed • Large variation in usage limits ability to identify savings • Sample usually not randomly determined • Isolating non-participants difficult in some types of programs (*e.g.*, large C&I who might have taken part in prior years) • Unobserved influences—sometimes reflecting the decisions of evaluators on which variables to include or exclude—may bias estimates • Does not include trade ally effects • Cannot disaggregate free ridership or spillover estimates, although they are sometimes embedded within the net savings estimates. • Large customers can overly bias results • Large C&I: Energy use data available only at building level • Can be expensive • In some cases measures adjusted gross savings, not net savings	Haeri *et al.* (2005) on transition commercial retrofit programs for the Energy Trust of Oregon

[47] This is a very important advantage, which by itself can outweigh the disadvantages

Method	Definition	Advantages	Disadvantages	Example*
Cross-Sectional Studies	Comparisons of market share of targeted technologies or behaviors between a baseline area not served by the program and the area served by the program; baseline can be actual geographic comparison area or a modeled geographic comparison area made to resemble the program in all respects except the presence of the program	• Can estimate net energy impacts for program where participation is not well defined • Compares actual behaviors • Addresses trends in the entire market for equipment	• Availability and quality of sales and shipment data • Even if available, data acquisition may be expensive and gaps can be extremely misleading • May be difficult to determine the appropriateness of a comparison area • Can be expensive	Winch and Talerico (2008) on CFLs in Wisconsin; data are cross-sectional in nature but presented for multiple years
Structured Expert Judgment	Panels of experts on relevant technology, infrastructure systems, markets, and political environments asked to estimate baseline market share for measure or behavior and, in some cases, forecast market share with and w/o program in place; Delphi process is the most widely known technique	• Independent from evaluators—based on the professional judgment of experts • Useful for programs with diverse and complex end-uses or practices	• Possible incorrect assignment of free ridership values by panelists • Cannot assign statistical precision	NMR *et al.* (2004) on various ENERGY STAR qualified appliances; includes forecasts to future years

Method	Definition	Advantages	Disadvantages	Example*
Self-Reported Approaches	Relies on market actors' own description of program influence on purchase (end users) or promotion (suppliers) decisions to characterize the program's effect; based on surveys of participants and non-participants to estimate free ridership and/or spillover among other adjustment factors to gross savings	• Simple and less expensive • Flexible • Direct information on behaviors • Can be used to look at trends over time if used consistently	• Potential non-response bias • Poor reliability and validity of some survey questions and, therefore, responses • Respondents do not always recognize operative reason for their adoption of measure • Program influence may be unknown to respondent when word of mouth or advertising sparks their participation • Possible incorrect assignment of free ridership values by evaluators • Often conducted months after decision was made, affecting recall	Research Into Action and Cadmus (2009) on the Energy Trust of Oregon's Existing Buildings (formerly known as the Building Efficiency) program.
Enhanced-Self Reported Approaches	Complements the self-report methods with surveys of vendors, trade allies, review of program documentation and market data, *etc.*; data triangulated to estimate NTG	• Same as regular self-report • Better than regular self-reports for estimating spillover	• Same as regular self-report • Can be expensive	Schauer (2010) on the Home Performance with ENERGY STAR program in Wisconsin.

Method	Definition	Advantages	Disadvantages	Example*
Macroeconomic Modeling	Uses state or national economic data at sector level (*i.e.* residential, commercial, industrial) to assess the extent to which markets for energy-efficient products and services have been affected by programs; data include energy prices, per capita income, heating / cooling degree days, technology change data, and commitment to energy-efficiency policies	• Estimates net effects of all programs cumulatively	• Methods not fully developed • Likely involve greater error than bottom-up program-specific methods	M. Horowitz (2008) applies this method nationwide and across sectors

* Some of the examples include combinations of multiple methods. For example, some self-report studies are also cross sectional in nature.

3.1 Reliability, Validity, and Bias in Estimates

By far the most common concerns raised in the literature and the interviews regarding most approaches to estimating net savings are bias, reliability, and validity, with construct validity being a particular concern (Megdal *et al*. 2009). Although defined in Section 1.3, we believe the definitions are worth repeating here:

> **Bias:** The extent to which a measurement or a sampling or analytic method systematically underestimates or overestimates a value.
>
> **Construct Validity:** "the degree to which an experimentally-determined definition matches the theoretical definition." (Jonas 2005) [See also *validity* below]
>
> **Reliability:** The quality of a measurement process that would produce similar results on: (1) repeated observations of the same condition or event; or (2) multiple observations of the same condition or event by different observers.
>
> **Validity**: "the extent to which [the measure] measures what it is intended to measure." (Rossi *et al* 2004:220).

Although researchers estimating net savings struggle with each of these concerns, the issue of validity—including construct validity—is the most critical because estimating net savings requires "measuring" the counterfactual as a baseline. By definition, however, the counterfactual never happened, and therefore cannot be directly measured. Therefore, we have to "construct" what would have happened in the absence of the programs, and cannot be certain that the methods truly captured this fictional condition.

In the sections that follow, we elaborate on some of the key concerns with measuring net savings overall and with some of the individual approaches used to estimate net savings. The project team believed an elaboration on these issues of greatest concern to the energy efficiency community in general—and Forum and Steering Committee specifically—was warranted given the feedback we received on earlier drafts of this paper. However, we do not present this section as an attack on any of the methods. Despite some of their shortcomings, each of the methods has its appropriate uses in the toolbox evaluators use to measure net savings.

3.1.1 Difficulty in Measuring Cumulative Program Effects over Time

One challenge for measuring net savings is that evaluators are expected to develop an estimate for a limited period of time, such as the previous program year (*e.g.* 2009) or previous program cycle (*e.g.* 2006-2009). Yet a given energy efficiency program may have been operating for years prior to the program period under question; moreover, the same program administrator may have been running other efficiency programs prior to—and concurrent with—the program and period for which net savings is now being estimated. Even if net savings estimates have been developed for this prior program activity, it is likely that the earlier program activity exhibited cumulative spillover effects during the more recent program cycle. Thus, participants who today

appear not to have been directly influenced by the program period under question may have been influenced by the program in prior years. The underestimation of cumulative effects creates concerns for both validity and reliability. Spillover from program activity in previous years may be identified as free ridership in the current program year, such that the estimate does not capture what it was intended to measure. Furthermore, if one assumes that cumulative effects are always positive, then the failure to account for them would systematically bias estimates of net program savings downward, both by failing to account for post-evaluation spillover in previous calculations of net savings and by classifying such spillover in the "current' evaluation as free ridership. [48]

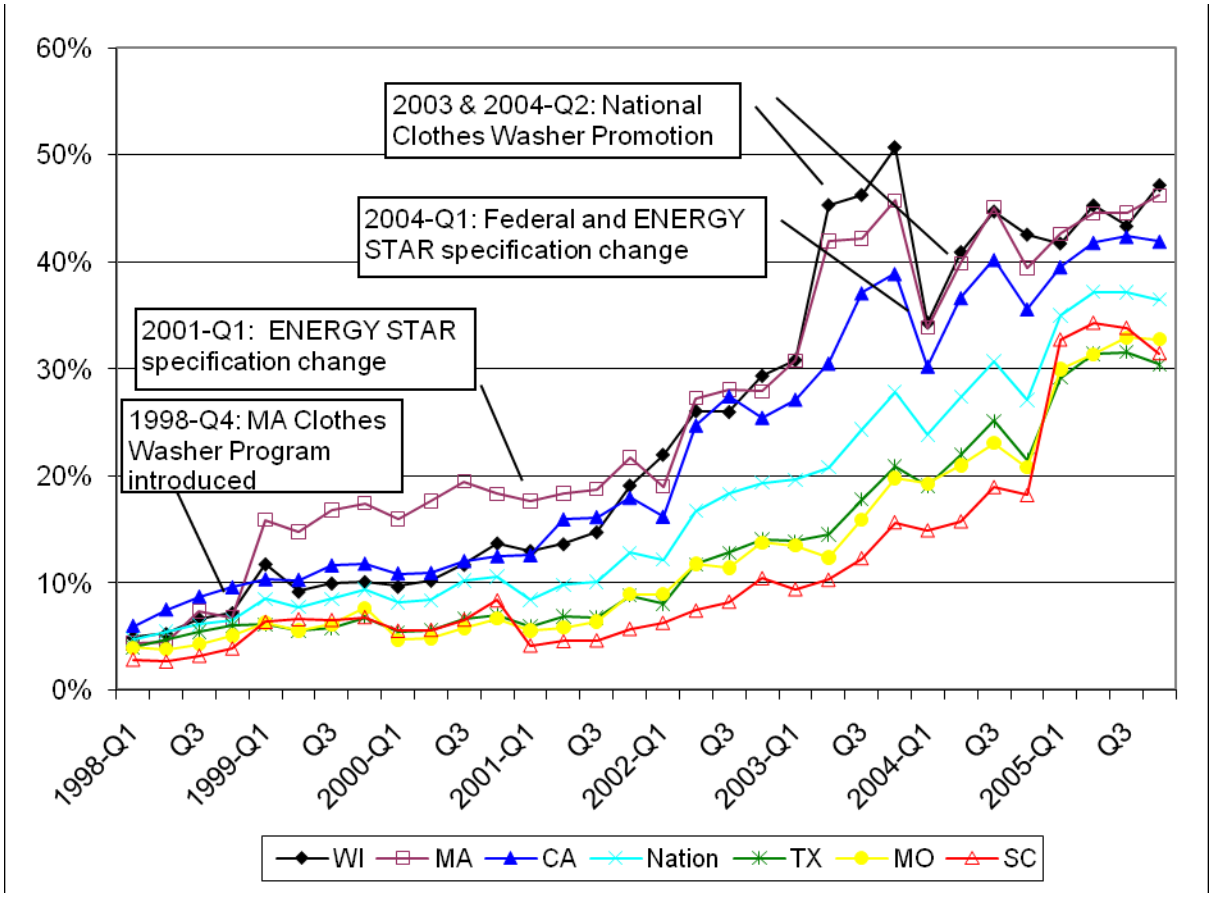
Rarely do net savings evaluations attempt to measure the impact of prior programs, and even when they do, capturing those impacts sometimes eludes evaluators (see NMR 2010). An evaluation of the Massachusetts ENERGY STAR Appliances Program is among the few studies that have successfully included cumulative program effects in estimates of net savings (NMR and Shel Feldman Consulting 2005; results also reported in various conference papers and presentations by Hoefgen, Wilson-Wright, Feldman, and Li, 2004 and 2005). The evaluation relied on a regression model that included current program activity and the cumulative effect of prior market share, together with demographic and economic factors to explain current market share of ENERGY STAR appliances. The cumulative effect variable was a key predictor of current ENERGY STAR market share for all appliances except dishwashers, which, at that time, had a very high market share nationwide in both program and non-program areas. Moreover, the evaluation also used canonical correlation to demonstrate that program support for clothes washers also boosted sales of ENERGY STAR refrigerators and room air conditioners even when program administrators did not offer incentives for these last two appliances.

It was possible to identify the cumulative program effect for the Massachusetts ENERGY STAR Appliances program because the Department of Energy (DOE) and its contractor D&R International (DRI) had been tracking the market shares of ENERGY STAR appliances by state since 1998. Although these market share data were not perfect—only four national retail chains contributed data and they did so inconsistently over the years—they provided a longitudinal data set that made possible the identification of cumulative market effects not only in Massachusetts but also in other states with active appliances programs (see tracking data in Figure 3–1, with Wisconsin, Massachusetts, and California representing active program states through 2005 and the other three states representing non-active states during the same period). Importantly, the DOE and DRI have recently stopped tracking these data, in part because of the difficulties they had in securing the necessary sales data from the four national retail chains—all of which are ENERGY STAR partners and have received rewards from ENERGY STAR for their efforts to promote efficient products. Without such data—and similar data for other types of programs—

[48] Regulators have the interests of ratepayers in mind when they require the estimation of impacts for a given program year; they want to know whether continued expenditures are required for savings to continue. However, the focus on a single year does not allow recognition of how current expenditures may produce savings in future years.

future efforts to measure net savings for a single program year but especially cumulatively over time and across programs and products will be severely hampered.

Figure 3–1: Quarterly ENERGY STAR Clothes Washer Penetration, 1998 to 2005
(as reported by DRI)



3.1.2 Failure to Account for Synergistic Impacts of the Portfolio of Programs

The second reason current approaches to net savings estimation create concern is that the single program for which net impacts is to be estimated is usually just one in a portfolio of programs as well as marketing, outreach, education, and training activities delivered by the program administrator. These multiple, diverse activities provide a broader context within which any single program operates. Net savings research focused on evaluating individual programs could potentially conclude that each activity had a low NTGR, perhaps leading some observers to conclude that all activities should be revised or discontinued because, based on the NTGR, most of the gross savings would have occurred in the absence of the program. Yet, it is possible—although not guaranteed—that the entire portfolio of programs and activities has synergistic effects not captured by measuring the net savings for each program and summing the results; the whole may be greater than the sum of its parts. To draw the conclusion that individual programs or activities with small NTGRs should be discontinued or revised in isolation does not recognize this possible portfolio-level synergy and could result in decreasing the net energy savings

achieved by the program administrator. This concern would be mainly about bias—the systematic undercounting of the true program savings resulting from a portfolio—although validity would also come into question as measuring net savings for each individual program or activity would not be measuring the intended overall effect—including spillover—of the program.

3.1.3 Concerns Raised by Cross-Cutting Methodological Approaches

The remaining concerns about bias, reliability, and validity are best discussed regarding three general approaches to measuring net program impacts in general and net savings specifically: 1) randomized controlled experiments, 2) quasi-experimental approaches, and 3) self-report approaches.

Randomized Controlled Experiments: It is generally accepted that randomized controlled experiments offer the strongest design for ensuring reliability and reducing the likelihood of bias. In randomized controlled experiments, individuals are randomly assigned to a treatment (program) or control (non-program) group, theoretically apportioning any factors that may produce bias evenly across the two groups. Measurements are taken from the two groups, ideally before but certainly after the treatment is applied, and the results are compared. These measurements are then usually analyzed quantitatively through such methods as difference of differences between the two groups or statistical modeling, including randomly assigned billing analyses (Allcott and Mullainathan 2010; see below for a discussion of the strengths and weaknesses of billing analysis and statistical modeling).[49] Moreover, because the procedure should be replicable, the approach should have a high degree of reliability which would be demonstrated by a separate group of evaluators achieving similar results using the same methods in a nearly identical situation. While this approach is the foundation of the experimental sciences, it is not immune to the problems of reliability, bias, and especially validity. During his debate with Cook, Scriven argued that individuals in the control groups tend to "drop out" of studies more frequently than those in treatment groups, leading to biased results which would also affect the reliability of the method in that the results would be difficult to replicate (Cook *et* al. 2010). Furthermore, the approach does not guarantee that the measurements truly capture the underlying construct of what would have happened in the absence of the program.

Quasi-experimental Designs: In broader evaluation circles, quasi-experimental designs are frequently considered to be the next best option to randomized experiments (Rossi *et al.* 2004; Howell and Yemane 2006). In these approaches, a treatment group is compared to a non-randomly selected comparison group. The control group may be matched to participants on key characteristics or the two groups may be matched on the aggregate characteristics of individuals within a geographic area (as in some recent CFL studies, *e.g.*, Winch and Talerico 2008; NMR, RLW Analytics, and Dorothy Conant 2008; Cadmus, KEMA, Itron, NMR, and A Goett 2010).

[49] Qualitative techniques can also be used in some situations (*e.g.* performing content analyses of interviews conducted with the treatment and control groups), but the methods and results must be replicable for strict adherence to the design of randomized controlled experiments.

Quasi-experimental designs are commonly used in energy efficiency evaluation, with some of the most common approaches including non-randomly assigned billing analyses (Haeri, Gage, Perussi, and Ogle 2005), differences in trends between two or more areas (Winch and Talerico 2008; Cadmus *et al*. 2010), and using statistical models to control for differences between program participants or areas and non-participating individuals or areas (NMR and Feldman 2005; Cadmus *et al*. 2010; NMR 2010).

Although we will not delve into a detailed accounting of the challenges of quasi-experimental designs for bias, reliability, and validity, a few observations are worth noting. Evaluators sometimes use billing analysis based on non-randomly assigned groups to estimate net savings. As mentioned in Table 3–1, the primary benefit of billing analysis—and it is an important one—is that it can provide statistically valid estimates of net savings under the right conditions. The concern, however, is when the conditions are not "right," and this happens frequently. Table 3–1 highlights the numerous limitations of the billing analysis approach; here we focus on three. First, billing data fluctuate a great deal for reasons often unknown to the evaluator. This is often treated as random variation in billing analysis, but it may be of comparable or greater size than the expected energy savings, leading to imprecise measurements. While the results may be reliable in a statistical sense, they could lack validity if the results do not truly capture net savings.

This leads to the second concern with billing analysis (and statistically modeling more generally—including that based on random assignment): the potential influence that variables not included in the model may have on energy savings. No statistical model can include every possible factor that may affect net savings. Researchers have to make educated decisions about which variables to include and which to exclude, and some potentially influencing factors may be completely unknown and never even considered as model inputs. If an excluded or overlooked variable actually has a strong effect on net savings, the resulting estimate will not accurately describe net savings—it will lack statistically validity and will most likely be biased as well.

The third factor relates to the fact that true, comparable non-participants may be difficult to find, especially for large commercial and industrial (C&I) programs. The pool of non-participants is typically sizable for residential programs, but this is not the case for large C&I programs. Evaluators have found that identifying comparable cases—especially those that have never previously participated in the program or previous ones—to be extremely difficult. In fact, as Ridge and his colleagues 2009 explain, this is one of the reasons many program administrators and evaluators have adopted the self-report approach over billing analysis to measure net savings. The inability to locate true, comparable non-participants would certainly lead to bias and validity concerns in the results.

Self-Report Methods: Most self-report methods of estimating net savings ask respondents a series of questions designed to determine if they can be classified as free riders or spillover, or,

in some jurisdictions, as exhibiting rebound.[50] The surveys typically include a series of questions because participant decision making is complex and a single-question approach can give misleading findings (Saxonis 1989; Windell and Peters 1994; Megdal *et al*. 2009). Evaluators develop algorithms to code the responses to these questions into a free ridership or spillover score or a NTGR (Rathbun *et al.* 2003; California Public Utilities Commission 2007; Winch *et al.* 2008).

The self-reported approach is currently the most frequently "maligned" (to borrow a phrase from Ridge *et al.* 2009) approach to measuring net savings. The most contentious aspects of the approach involve the questions themselves and how they are used to estimate net savings as well as the underlying psychology of how respondents answer the questions.

One of the criticisms of the self-report approach is that it is susceptible to evaluator subjectivity, which also affects reliability, validity, and bias.[51] Subjectivity arises due to two factors: 1) the scores assigned to partial free ridership, and 2) the resolution of inconsistent responses. Evaluations differ as to the increments they use for degrees of partial free ridership, such as whether the lowest increment of possible free ridership is 10% free rider, 25%, or some other number. This subjectivity in the quantification of degree of partial free ridership makes it problematic to compare evaluation results across different programs, and raise questions about both the reliability and validity of the method.

The second factor related to subjectivity applies to all surveys that use multiple questions to measure free ridership or spillover. At times, respondents will give inconsistent answers to the questions, making it even more difficult to classify their actions as free ridership or spillover. In such situations, some methods require the interviewer to determine whether inconsistent answers have been given and, if so, to ask additional questions. This approach can result in differences in the probing done by different interviewers (Rathbun *et al*. 2003) or may alternatively exacerbate the frustration by some respondents who perceive that they are answering the same question multiple times. In the analysis of inconsistent answers, some evaluators seek to increase reliability by having multiple raters collaborate, ensuring inter-rater reliability (Maxwell, Gregoire, Meissner, and Megdal 2009). While this approach reduces subjectivity, there remains the possibility that a different team of evaluators would generate different results. Megdal and her colleagues (2009) addressed inconsistency in responses through follow-up in-depth interviews through which they were often able to explain many of the inconsistencies based on the unique situation of each project. It must be noted that this research focused on large C&I customers and the sample size was relatively small; such a follow-up approach is not always

[50] Some self-report methods for measuring net savings—especially for upstream programs—ask respondents to estimate how many of a product they purchased and when they purchased the products. Evaluators usually analyze these data using a quasi-experimental method (Winch and Talerico 2008; NMR 2010).

[51] Of course, subjectivity is not unique to the self-report approach, as it also relates to the variables included in statistical models, the factors on which non-participants are matched to participants, and the evidence considered in such approaches as historical tracing and structured judgment (see Table 3–1).

feasible, especially for residential programs with thousands of participants in a program year or cycle.

Additional debates relate to the fact that most self-report approaches rely on a series of questions to determine free ridership and spillover in order to capture different aspects of these two concepts and to estimate partial free ridership. The energy efficiency community considers the use of multiple questions to estimate full and partial free ridership and spillover to be a best practice. The need for multiple survey or interview questions directly results from the difficulty of measuring the underlying construct of what would have happened in the absence of the program as well as the various ways in which a participant can be a free rider or exhibit spillover. Debates develop, however, over the number of questions to ask, which questions to ask, and how to address the seemingly contradictory results that typically manifest when using multiple questions to measure the same concept.

Those advocating for longer batteries assert that the numerous and detailed questions, such as those found in the California Self-Report Approach, can tease out the actual reasons for behavior and are needed capture the nuances of free ridership and spillover; without the detailed questioning, evaluators will not be able to determine with any degree of certainty if the respondent is a free rider or spillover (Ridge *et al.* 2009; Megdal *et al*. 2009). Although the surveys can be lengthy to answer and seem repetitive and tedious to some respondents, these advocates argue that the use of highly trained evaluators to design and deliver the survey will lead to successful results (Megdal *et al.* 2009). Other researchers, however, argue the opposite. They assert that the lengthy battery of questions attempts to capture a level of detail best collected during in-depth interviewing and not telephone surveying as respondents cannot discern the nuances among the questions, often saying, "I just answered this," and becoming audibly angry and frustrated at the length of the survey. Although we are not aware of any research that has concluded that the length of the battery of questions affects the validity of the results, survey implementers report that longer surveys have lower response rates due to a greater number of mid-survey break-offs from respondents who become frustrated with the length and seemingly redundant questioning (Personal communication; Holbrook, Krosnick, and Pfent 2008). Because of this, other evaluators now advocate for a shorter battery of questions delivered closer to the time of decision making—that is, soon after participation if measuring free ridership and later if measuring spillover (Peters and Bliss 2008). By timing the survey closer to the point of decision making, respondents more accurately recall why they took an action without the need for a lengthy battery of questions to tease out these reasons. This approach may require the delivery of two surveys—one for free ridership and the other for spillover at a later time—but the shorter survey length allows this from both a customer service and budgetary standpoint.

Another source of contention regarding self-report approaches relates to the underlying psychology that drives the decision not only to adopt the measure and take part in the program, but also on how to respond to the questions in a survey. For example, drawing on the work of Mohr (1995, 1996), Ridge and his colleagues (2009) argue that individuals can identify the

reasons for their actions and report these to researchers. Mohr (1995, 1996) argues that research into behavior can uncover what he calls the *operative reason* for a behavior—the one factor or force among many that actually leads an individual to act. At times, an individual can accurately name the operative reason for taking an action, particularly when the reason produces a strong emotional response in the actor. According to Mohr (1996), more frequently, however, these operative reasons remain unknown to the individual, who will cite *thoughts* about taking a particular behavior as the reason for doing so, even though these thoughts existed prior to the action and had not previously produced the behavior in question. To draw on an energy efficiency example, a respondent to a self-report survey may report that she would have installed the measure without the program in order to protect the environment, but, in actuality, this thought had previously existed. It was not until she learned about the incentive that she actually followed through on her intention to install the measure. The incentive serves as the operative reason for adopting the measure, even if the participant is not fully aware of its impact on her behavior (see also Peters and McRae 2008). In fact, Mohr (1996) explicitly singles out the adoption of innovative technologies (which include energy efficient products) as an area in which identifying the causes of behavior is particularly difficult. Citing research by Rogers and Shoemaker (1971) on the predictors of innovation, Mohr explains that the research on predictors of innovation:

> *showed a substantial spread of results on each predictor variable—sometimes even ranging from significantly positive as a predictor to significantly negative. Given that innovation is governed by encounters [a convergence of probabilistic circumstances], this result is unsurprising in principle. What many of us seem to hope for is that,* in practice, *we will nevertheless find predictors so endowed with affect and associations that they win out in almost every actual encounter…. Unfortunately, our experience with discovering good, strong, important, and broadly applicable cause-and-effect relations is negative—we have not made any such discoveries* (1996:111-112, emphasis in original).

The parallel exists in net impact evaluation. Estimates of free ridership, spillover, and NTGR often appear to be "all over the place" and demonstrate no recognizable patterns or consistency. Program administrators and regulators continue to search for an evaluation approach that will demonstrate unequivocally that the incentives and other program activities provided that "good, strong, important, and broadly applicable cause-and effect relationship," but such a finding eludes the energy efficiency community. If Mohr is correct, the challenge of finding a strong and consistent connection between program activity and measure installation may be because the adoption of efficiency technology requires the convergence of multiple thoughts and circumstances; the incentive or program activity becomes a vital part of this encounter, but not in such a way that the individual participant can always recognize its effect beyond those of the other influences on the behavior.

The study of human psychology raises other concerns about self-report approaches related to why respondents give particular answers to surveys. One issue frequently raised is that of the

"socially desirable" response. Historically, many members of the energy efficiency community argued that survey respondents would report that the program influenced their behavior to please the evaluators and protect the existence of the program (Peters and McRae 2008). More recently, however, commentators have instead concluded that respondents say that they would have adopted an energy efficient technology or behavior without the program because they want to appear to be concerned about energy efficiency and the environment (Adams, Soumerai, Lomas, and Ross-Degnan 1999; Peters and McRae 2008). Giving this answer avoids the psychologically challenging state of cognitive dissonance that would exist if the person admitted that their actions did not reflect their self-image as the type of person who behave in such a manner; moreover, the respondent may be so pleased with the outcome of participating that they are convinced they would have taken the action anyway, and by doing so, they have become "just that type of person" who takes actions to save energy and protect the environment (Peters and McRae 2008; In-depth interviews).[52] While the energy efficiency community attempts to improve the battery of questions and methods used to assess net savings via self-reported approaches (Saxonis 2007a; Dyson and Goldberg 2007), some commentators do not believe incremental—or substantial—changes to the battery of questions or delivery methods will resolve the problems with self-reported free ridership or spillover because the difficulties have little to do with the methods and much more to do with human psychology (Peters and McRae 2008).

Despite criticisms, the self-report approach is among the most commonly used approaches to measuring net savings for four very important reasons:

1. The approach is flexible, allowing the ability to tailor questions to the particular nature of the program design, implementation method, and product characteristics.
2. The approach often costs less to implement than many other approaches, and can provide additional information useful to process and impact evaluation such as program awareness, satisfaction, and demographic data can be collected at the same time, leveraging evaluation resources.
3. The approach yields estimates of free ridership, participant spillover, and the NTGR without the need for a non-participant comparison group, an important characteristic when comparable non-participants do not exist or are difficult to identify (Ridge *et al*. 2009).
4. The approach makes it is easier to develop a consistent set of questions and specifications of varying levels of partial free ridership and spillover that can be used over time to see how the estimates are changing.

[52] Peters and McRae (2008) explain that many discussions in the energy efficiency community regarding the socially desirable response assume this would lead the respondent to say the program caused the action in order to please the interviewer who has called on behalf of the program. However, they argue—and interviewees and other NMR Team members concur—that it is increasingly likely the response tends the other way, with the respondent saying she would have taken the action without the program because it is socially desirable to be the type of person who regularly acts to save the environment, not one who needs a rebate to do so.

For these reasons, the self-report approach can and should remain an important tool for estimating net savings. Like the other approaches, it has its shortcomings that must be recognized and, when possible, guarded against, but the energy efficiency community should not be throwing tools out of the methodological toolbox; instead, we should be striving to improve current methods, introduce new approaches, and secure higher quality data in order to move forward the reliable, valid, and unbiased estimation of net savings.

3.2 Logistical Concerns

The second major set of concerns about net estimation methods is logistical in nature. In this category are such things as the available budget; the timeline of the evaluation; the availability of data; the quality of data; the size and importance of the program; the likelihood of free ridership; spillover, and market effects; and the results required by regulators and other primary audiences for net savings estimates. Each of these concerns influences the choice of methods for estimating net savings, and those choices have concomitant effects on reliability, validity, and bias. As Rossi *et al.* (2004) explains, evaluators often—and sometimes should—go with the method that is "good enough." The "perfect" approach may not be feasible or even advisable in some situations.[53]

[53] For example, evaluation resources should not be expended on estimating market effects of a small pilot program.

4 Net Savings in Relation to Current and Evolving Policy Needs

Section 2.1 above discusses the degree to which the audiences interested in net savings estimates have recently grown and diversified, in large part because of newly created programs and expanded savings goals resulting from ambitious state, regional, and national energy, climate change, and economic recovery policies. These policies come not only with increased funds for implementing and evaluating energy efficiency programs, but also with increased scrutiny of program performance, as well as requirements for which outcomes are reported and in what manner. In some cases, energy or demand savings—be they gross or net—represent the primary outcome of interest (*e.g.* for the forward capacity markets), but in others, the calculation of energy savings is a critical step in the calculation of the targeted outcome, such as reductions in greenhouse gas emissions.

The energy efficiency community had grown accustomed to—but by no means complacent about—debates regarding the advantages and disadvantages of various methods and definitions of net savings, or even whether it is necessary to calculate net savings at all. The community had consistently made strides in improving the conceptualization and measurement of net savings, but it had largely accepted, sometimes reluctantly, that reliability, validity, and bias concerns will persist and jurisdictions will continue to differ in their approaches to net savings. Newer audiences for energy savings estimates—whether gross or net remains unclear—have "stirred the pot." They are expecting energy savings estimates to provide them with the vital information they need to assess the performance of their own programs and policies. With this expectation come new requirements for energy efficiency program administrators regarding the types and quality of information they must submit to the expanded audiences. In this section, we discuss the degree to which the current approaches to net savings estimation meet existing policy needs, before turning to a similar discussion of the ability of current approaches to meet evolving policy needs. We pay particular attention to the needs of air regulation community, as many commentators believe this group will be one of the primary audiences for energy savings estimates in the near future.

4.1 Meeting Current Policy Needs

Commentators had mixed opinions as to whether net savings approaches meet current policy needs. Often, the same individual would argue that the current approaches meet some policy needs but not others. Generally, energy regulators were more likely to state that the approaches meet current policy needs, while program administrators and net savings experts voiced greater skepticism on this topic, but even this generality involved exceptions, as discussed below.

4.1.1 Net Savings Approaches Do Meet Current Policy Needs

Those commentators saying that net savings approaches do meet current policy needs asserted that net savings estimates provide a critical piece of information about how well the program is performing. Net savings estimates help program administrators, planners, and regulators understand the strengths and weaknesses of the program design and decide when to stay the course, revise a program, or discontinue it. Many of these authors and interviewees concede that aspects of impact evaluation other than net savings may also elucidate broader program benefits, which should be taken into account when assessing the overall program impact. Yet, many of these sources still asserted that net savings will and should remain one of the most important considerations in assessing programs, largely because it is the responsibility of program administrators and regulators to ensure that ratepayer and taxpayer funds are spent wisely and achieve the greatest return on investment in terms of energy savings.

4.1.2 Net Savings Approaches Do Not Meet Current Policy Needs

Commentators offered two explanations for why they did not believe that approaches to net savings meet current policy needs. The first explanation rested on skepticism regarding the quality of existing methods and, therefore, the results. As one program administrator stated numerous times, "We're underestimating our effects." Many of these individuals argued that approaches to net savings could meet current policy needs if they could more accurately estimate free ridership and spillover, as well as cumulative effects and market effects, including not only the market effects of explicit market transformation programs but also of other programs that potentially have market-level impacts.

The second explanation was that net savings is too narrow a focus and fails to recognize the broader context in which current programs (and future ones too) operate. The commentators argued that it is short-sighted to spend resources trying to determine net savings. They would prefer to focus instead on formative research on how to change behavior and engage the customer in order to improve programs and achieve deeper savings and emissions reductions. Programs, in this view, should be evaluated in terms of how well they engage the customer and should involve measures like extent of behavior change and market transformation. This group of commentators held that the focus on minimizing free ridership is counterproductive and greater savings could be achieved if the focus were on maximizing outreach and not on avoiding outreach to the "wrong people" (*i.e.*, free riders). It is important to note that this group thought that the current policy focus on net savings—more specifically free ridership—impairs planning and innovation that will allow programs to meet evolving policy needs.

4.2 Meeting Evolving Policy Needs

We asked interviewees to discuss the evolving policy needs regarding net savings, particularly those focused on expanded energy savings goals, reduced greenhouse gas emissions, and jobs creation. Policies of interest included energy savings goals at the state level, and the federal

American Recovery and Reinvestment Act (ARRA) and similar state stimulus funds slotted for energy efficiency programs and green jobs. Moreover, we also inquired about the impact of the Regional Greenhouse Gas Initiative (RGGI) in the Northeast, which currently does not have regulatory overview power, but it is possible, and some would say likely, that there will be directives at the federal level that will set forth directives concerning how to treat energy savings in estimating emissions reductions. The take-away message can be summarized as follows: 1) Things are changing and fast, 2) Programs will need to adopt new design and implementation approaches to achieve the ambitious goals that have been set, and 3) Program evaluation will need to make significant adjustments in response to these changes, but 4) We are not quite sure what these changes will entail or what will be needed, so we are uncertain about what direction to take. These circumstances contributed to differences regarding the degree to which interviewees said that current approaches to net savings would meet evolving policy needs.[54]

4.2.1 Reporting Needs for Legislators and Federal Regulators

The United States Department of Energy (DOE) and the Environmental Protection Agency (EPA) have long been involved in energy efficiency programs and promotions, such as the Low Income Heating Energy Assistance Program (LIHEAP), ENERGY STAR, and the *Change a Light* CFL campaign, to name a few examples. However, the recent passage of ARRA brought with it a host of energy-related funding for appliance turn-in programs, state environmental programs, energy efficiency community block grants, and green jobs programs. Given the large amounts of funding being awarded to the states through ARRA and administered by the DOE—as well as the close scrutiny of their use and impact by the federal government and watchdog groups—program administrators, state-level energy regulators, and evaluators are scrambling to understand just what it is they will have to report to the DOE. In addition, some state legislatures in the Northeast have recently passed ambitious energy use and greenhouse gas reduction targets, and, again, the scrutiny being placed on funds allocated to meet these goals is high. In fact, program administrators in one state informed project team members that the state attorney general had weighed in on the type of evidence that should be allowed in determining program impacts.

Given that these programs—and the level of scrutiny they receive—are quite new and have goals that extend beyond saving energy for its own sake, our extremely knowledgeable interviewees disagreed on whether current net savings practices and approaches were adequate for responding to recently evolving policies. One state-level energy regulator said the DOE was most interested in gross savings and would use those in its cost effectiveness tests for ARRA-funded programs. Another interviewee disagreed and argued that the DOE was usually interested in net savings, but also stated, “It will be interesting to see how net savings is handled for different programs

[54] We do not repeat here the differences of opinion regarding whether energy regulators should be most concerned with gross savings, net savings, or broader indicators of program success. The points made by interviewees regarding current policy needs of energy regulators apply also to what are perceived to be the future policy needs of energy regulators.

like ARRA- and RGGI-funded programs" in which job creation and greenhouse gas reductions, respectively, and not energy savings, are the targeted goals. Please note that our interviews preceded webinars sponsored by the DOE to clarify evaluation and reporting guidelines for ARRA, in which presenters stated that the DOE was indeed interested primarily in net savings (Hall 2010). Had we conducted them after the webinar, we may not have observed the confusion reported here.

4.2.2 Energy Savings Estimation Needs for the Air Regulation Community

Differences in opinions on whether current approaches to measuring energy savings—whether adjusted gross or net remains unclear—meet evolving policy needs were perhaps most notable regarding air regulation. While RGGI has provided funds to existing energy efficiency programs (or new programs very similar to existing ones), thereby supporting current evaluation approaches, there is the potential for a federal greenhouse gas regulatory program to exist in the future. Multiple commentators believe it makes sense for the energy efficiency community to plan for this eventuality, not by designing new programs, but by working with the air regulation community to develop evaluation and reporting approaches that support both communities. Like their energy efficiency counterparts, air regulators would like reassurance that emissions estimates resulting from efficiency programs are valid and measurable. This section summarizes some of the main topics raised regarding energy savings as it relates to greenhouse gas reductions.

Adjusted Gross Savings vs. Net Savings: The core issue of concern relates to whether the air regulation community will be required to base savings on adjusted gross savings or net savings. Some interviewees asserted that air regulators will likely be concerned only with gross savings, with one stating, "My guess is they will need total changes in energy use in society, not just net effects of programs, but I don't know how they're going to measure that." A different energy regulator suggested that net savings will be an important to air regulators, but air regulators had not yet become concerned with the issue. At this time, he argued, air regulators were focused on securing quality monitoring and verification in general, and he believed their concern with net savings would follow. The air regulators we interviewed stated that they needed to demonstrate savings above a baseline, but they differed in whether they believed the necessary documentation of change from the baseline would reflect gross or net savings, perhaps because some of the air regulators remained unclear on the distinctions between the two types of savings.

The confusion over whether the air regulation community will need net or gross savings stems from the requirement that air regulators must be able to prove that emissions reductions have occurred. One air regulator strongly supported the use of adjusted gross savings to estimate emissions because he argued that adjusted gross savings more closely align with the need to prove that savings occurred through the use of reliable and valid methods that involved actual measurements of energy savings, not estimation of a counterfactual. He held this view because he concluded that the requirements set forth in State Implementation Plans (SIPs) for greenhouse

gas reductions would not allow for the uncertainty in net savings estimates.[55] The air regulator arguing for the use of adjusted gross savings did leave open the possibility that advanced metering or other measurement-based approaches may provide an opportunity to claim net savings if evaluators can develop methods of converting the information from such measurement devices into net savings in a way that meets the requirements of SIPs.

Another complicating factor regarding the use of adjusted gross savings or net savings to estimate emissions reductions is the operational meaning of the term additionality. As defined in the Forum's *Glossary of Term* "additionality" is "a criterion that says that avoided emissions should be recognized only for project activities or programs that would not have 'happened anyway' in relation to a baseline estimate of project activity and associated with emissions reductions." Yet, what additionality means in practice is the subject of debate. While some commentators interpret the additionality requirement to necessitate estimates of net emissions reductions, others interpret additionality as a requirement not to double-count reductions by attributing the same action to two different interventions.

Measurement of Baselines: As mentioned above, air regulators will be determining emissions reduction compared to a baseline. One energy efficiency expert, however, voiced concern about the nature of these baselines. He suggested that air regulators "might define energy savings as observed changes in energy use in society, compared with their models or forecasts of total projected energy use." He stated that energy savings should not be double-counted but was also concerned that expected savings from energy efficiency programs were not considered as part of the baseline forecast. "This approach will work if their models don't include projected effects of programs and codes and standards." In other words, the models used to build baselines should not include the effects of existing or expected energy efficiency programs or regular updates to energy code. Instead, those programs and codes should get counted as reductions against the baseline, not as the "business as usual" scenario. The interviewee said he found it difficult to find information about the inputs into models and asserted that more work was needed to understand the baseline forecasts to ensure accurate and fair counting of reductions.

Linking Energy Savings to Power Plants: Another consideration revealed by our research is that air regulators are just beginning to understand the information they will need to convert energy savings to emissions reductions. One air regulator explained that, "To start with, we need megawatt and therms savings. But … some of the analysis that is already going on looks at differences in what avoided emissions might be based on time of day and time of year." Air regulators would prefer—and perhaps one day require—efficiency program administrators to provide energy savings estimates by time of day and time of year, because less efficient power plants with greater greenhouse gas emissions are the primary concern, and their use varies

[55] An additional concern with the need to prove savings is that the ambitious goals for energy savings and emissions reductions currently in place in multiple jurisdictions in the Northeast are expected to be achieved through multiple programs, including behavioral ones and others for which it may be challenging to measure gross savings, let alone net savings. This raises the question of whether such programs will be penalized or their savings not claimed because they cannot prove their more difficult-to-measure savings.

geographically and by time period. Thus, in some locations, measures that save energy at base load are more important for emissions reductions, while in other places measures that save energy at peak are more important—it all depends on the generation mix. Energy savings estimates by themselves would not allow the accurate estimation of emissions reductions without specifying when those savings occur. Although load shapes and coincidence factors are used to estimate demand reduction (which is not the subject of this paper), typically the energy efficiency community does not estimate energy savings by time of use. As one energy regulator noted, "The biggest problem for [air regulators] is that the way savings are currently recorded doesn't help them get at the things they are most interested in knowing." For this reason, he argued that the energy efficiency air regulation communities would have to think creatively about how to estimate emissions reduction from energy savings. Similar to the air regulators, he explained that the greatest reductions occur over the peak summer period, but further noted that summer peak events usually occur over a multi-day period of high temperatures. He did not believe that peak coincidence factors provided the nuanced information air regulators need, calling these coincidence factors "blunt instruments" that did not capture the importance of the multi-day events that lead to the peak. Instead, it is likely that that air regulators will have to understand savings (most likely adjusted gross savings) for each of the 8,760 hours in the year (see WRI 2007). Note that system planners, who are not currently considered an audience for net savings estimates in the Northeast due to their reliance on adjusted gross savings, will need to be a part of determining the best way to link adjusted gross or net savings to the power plants operating at the time when the savings occur.

In summary, approaches for translating energy savings into air emissions reductions are just now being developed, and "These questions are thorny. There are so many different parts, it's a moving target." At this time within the energy efficiency community, there is no policy requiring the development of energy saving estimation methods that also meet the needs of the air regulation community. However, some commentators said that the "writing is on the wall" that such policy is forthcoming. These individuals asserted that the needs of the energy efficiency, the systems planning, and the air regulation communities would best be met if these methods were developed in partnership, rather than having the air regulators decide the methods and the energy efficiency community and systems planners react to them. It is the understanding of the project team that the EPA is in the process of developing guidelines for how to claim emissions reductions from energy efficiency programs, which may help to provide clarity on the issues raised here.

4.3 Attribution in the Face of Multiple Programs and Policies

As mentioned earlier, attribution refers to the practice of determining what impacts are *caused* by a specific program or portfolio. Most commentators held that the estimation of net energy savings—what portion of gross savings a specific program has brought about—always involves determining attribution, but attribution analysis does not always produce estimates of net

savings. In other words, estimating free ridership and spillover *can* be a part of attribution analysis, but it need not be. Increasingly, the energy efficiency community is focusing a great deal of attention on one aspect of attribution: isolating the impact of the program from the impacts of other programs, messages, and contexts that could also influence the same or similar energy saving actions (Peters and McRae 2008; Skumatz 2009; and Mahone and Hall 2010).[56] This focus reflects the perception of some commentators that the complex web of programs, messages, and influences that encourage individuals to adopt energy efficiency measures makes the task of estimating net savings more difficult than before—ultimately, they argue, reducing the accuracy of recent net savings estimates. As ELCON (2008) argues, "Isolating the net savings of an energy efficiency program often has to deal with confounding market effects that exist, for example, when other publicly funded or private sector programs are targeting the same participant audience with the same, similar, or complementary energy efficiency measure."

Attribution was an important concern to interviewees. A few interviewees said that sorting out the issue of attribution was more important than that of measuring net savings in terms of free ridership and spillover. One air regulator stated, "The question of attribution is going to be the most difficult because of overlapping efforts." Two industry experts on net savings agreed, adding that we have to avoid double counting of savings. A program administrator staff evaluator shared a similar opinion, "A major complicating factor is that now we've got new funding sources, ARRA, different utilities [sponsoring programs], new standards and codes, programs that promote green jobs. How do we tease out effects attributable to each of these activities and funding sources, each of which has its own goals and ways of achieving savings?... Our current approaches will need to be modified." Other interviewees, however, argued that it is important not to lose sight of the overall goal of reducing energy use by focusing too narrowly on which program or source of funds gets the credit, because there could be synergistic effects among programs; the whole may be greater than the sum of the parts. There is a danger that the effectiveness of programs could be reduced if the energy efficiency community focuses program efforts only on what is most provable, not what is most effective.

The challenges of attribution have multiple roots, some psychological and others practical. As Peters and McRae (2008:5-224) note, "Attribution theory [in psychology] tells us that in the process of taking an energy efficiency action, the person acquires the attribute of being someone who takes such action." When asked if they would have taken the action without the program, the respondent says yes, because "I'm just that sort of a person." They further demonstrate that the problem is only exacerbated in this age of multiple programs and messaging when someone may become aware of a technology or behavior from any number of sources, but is only ultimately moved to act because the program facilitates their ability to do so (*e.g.* by providing an incentive to overcome a cost barrier, convincing an individual to translate their intention to action). Likewise, the actions of individuals are in reality caused by a confluence of multiple

[56] Economic concerns, habitat and species preservation, oil spills, wars in fossil fuel-producing areas, and even increases in gasoline prices are just a few of the broader circumstances that may lead people to save energy in an effort to save fossil fuels and electricity.

factors; attributing their actions to a single program or rebate would not only be difficult, but would also not accurately reflect the true decision making process (Bensch and Sulyma expressed these ideas as personal communication to Skumatz *et al*. 2009). As discussed in Section 3.1.3, Mohr (1996) would agree that the adoption of energy efficiency measures reflects a confluence of factors, but he would argue that an operative reason for the behavior *may* be able to be identified, although it could be difficult without in-depth exploration of potential causes of behavior.

Attribution concerns are not isolated to energy efficiency programs. DeGroff *et al*. (2010) discuss various challenges to attribution in public health (*e.g.*, complexity of problems, delayed outcomes, decentralized nature of program delivery, and unreliable data). Connell and Kubisch (1998) address similar issues in community development initiatives (*e.g*. programs offered at different levels of government targeting the same population and issues). Mayne (2001) suggests that attribution concerns persistently challenge all types of publicly funded programs precisely because they are complex and often address issues also addressed by other programs and organizations, not just the government.

Interviewees and the literature were not certain about how to take steps to resolve the attribution issue. One interviewee stressed the importance of using an appropriate baseline: "We have to make sure that, when we're developing a forecast model and see how it compares with observed savings in society, we know what is included." Similarly, an interviewee offered that "Money needs to go to developing reliable ways of measuring impacts of different sources of energy savings, reliable ways of attributing total savings to the right sources." In situations in which a program leverages funds from different sources, each of which wants to claim its own portion of net savings, the easiest solution may be to develop an overall estimate of net savings and then allocate the savings proportionately by how much money each source contributed to the program. This is the approach that DOE says it will accept as the minimum documentation of attribution in its ARRA guidelines, although more complex approaches will also be allowed (DOE 2010).

In more complex situations, some commentators argued, the estimation of net savings at the programmatic level should be dropped in favor of measuring savings (and emissions reductions) for the entire portfolio of programs offered by a program administrator or within a given jurisdiction or region, thereby better capturing the synergistic effects of the entire portfolio on net energy savings. Another idea identified in the broader evaluation literature—particularly the literature on public health, community development, and social services programming—is to use what is essentially a "weight of evidence" approach associated with historical tracing (see Section 3.1.3 and Table 3–1) (Mohr 1995, 1996; Connell and Kubisch 1998; Mayne 2001; DeGroff *et al*. 2010 also argue for this approach in other areas of program evaluation). Paired with a clear program theory and logic model, these approaches may be the best approach of attributing some outcome to a specific program. DeGroff and her colleagues (2010) also point to such practices as involving stakeholders when determining performance measures, piloting new

measures of performance in diverse situations, building flexibility into the measurement system to allow for site-specific contexts, and allow for more qualitative measures of outcomes, among other strategies, to facilitate attribution. Some of these suggestions could potentially be adapted to energy efficiency programs. Probabilistic approaches such as statistical modeling (Sulyma quoted in Skumatz *et al*. 2009:49) or structured judgment approaches such as Delphi panels may hold promise for confronting attribution concerns, although these approaches have their own strengths and weaknesses, as expressed in Section 3.[57]

[57] One commentator, speaking specifically about attribution, suggested that a company such as Proctor and Gamble (P&G) would not ask the question of whether their efforts or someone else's efforts led to the sale of their products. We concur with this statement, but argue that the analogy misses some key differences between a company selling consumer products and a program rebating a product manufactured—and promoted—by another company. Because P&G, to keep with the analogy, has the sales receipts—and money in the bank—it is unnecessary for them to make a claim about attribution. The sales are all the attribution they need. The same is not true for energy efficiency programs, which rebate products created by others. The message a consumer sends paying for a product is very different than accepting an incentive to offset the cost of something. The sale of an energy-efficient appliance accrues to the credit of the manufacturer, but not necessarily to the credit of an energy-efficiency program, at least without further proof. It is further the case that a company like P&G has ample resources at its disposal—and the willingness to use them—to assess the effectiveness of its promotional activities on sales net all the other factors that may also affect sales. They sponsor market research to understand how their campaigns affected sales, controlling for demographic, economic, and lifestyle factors. They recognize that sponsorship by celebrity or product placement on a population television show or movie can boost sales. They will still take credit for the sales and attribute the sales to their actions, but they understand that those sales are affected by outside forces and will assess the impact of those forces on their sales based on a baseline measurement or comparison group. If the energy efficiency community had the same control of the product and its promotion, resources, and access to sales data as does a company like P&G, it could potentially get to the point of making the same types of claims as P&G. But lacking such control, resources, and data, the energy efficiency community, the energy efficiency community will likely still need to determine attribution, at least to an extent, while accepting that the methods at our disposal will be subject to uncertainty.

5 Consistency in Net Savings Definitions and Methods

The project team explored interviewees' opinions about the possibility of adopting consistent definitions of and methods for measuring net savings throughout the Northeast. In some ways, the definition of net savings guides the methods; one must know which components are included in order to use each method. However, numerous methods can often provide the information required by the same definition or conceptualization of net savings, although the resulting estimates and their validity may differ. For this reason, net savings methods differ a great deal, even among jurisdictions using the same definition—or even within a single jurisdiction—let alone across jurisdictions with different definitions of net savings. Hence, methods for measuring net savings differ greatly even under situations in which the same definition is used.

5.1 Consistent Definitions of Net Savings

As we documented in Section 2.2 and Section 3, definitions of net savings and approaches to estimating them differ across jurisdictions and sometimes even across programs offered by the same program administrator or organization. In the literature review and in-depth interviews, the project team explored the idea of pursuing consistent definitions and methods of measuring net savings. The interviewees almost always supported the idea of a consistent definition of net savings, at least theoretically. The main reason the Northeast Region currently does not have consistent definitions, some argue, is that they do not have to. There has been no compelling public policy driver, such as legislation or interstate agreements, to stimulate the development of a consistent definition, so states have largely taken the path of least resistance and followed their own definitions. When it came to putting the theory into practice, the interviewees voiced some concern about what it would take to get the various jurisdictions to agree on a regional definition of net savings.

5.2 Consistent Approaches to Measuring Net Savings

There is a wider range of opinions on the issue of consistent methods. On one hand, it seems evident that consistent methods for measuring energy savings—be they adjusted gross or net—would be beneficial, particularly given regional and perhaps national goals for reduced greenhouse gas emissions whose achievement relies to a large extent on decreased energy use. To paraphrase multiple interviewees, air regulators need to know that a MWh saved in New Jersey means the same thing as a MWh saved in Connecticut in order to ensure that a program or administrator has met emissions reductions targets, and to facilitate trading credits across borders.[58] On the other hand, the diverse characteristics of programs, products, and participants may present challenges to efforts to institute consistent methods of estimating energy savings, depending on what constitutes that consistency. Air regulators generally were the strongest

[58] Air regulators may still have to apply different emissions factors to these estimate and account for when savings occurred (see Section 4.2)

advocates for consistency in estimation methods, regardless of whether they advocated for the use of adjusted gross or net savings estimates.[59] Net savings experts and program administrators—including some working in numerous jurisdictions—generally liked the idea of consistent methods for net savings estimation, but often raised concerns about doing so in practice, particularly given skepticism about the reliability and validity of methods and their ability to meet current or evolving policy needs. Energy regulators and their representatives voiced the greatest skepticism regarding the promotion and adoption of consistent methods, supporting their positions by pointing to the evolution of the diverse methods currently in use—namely that older methods have been found to be lacking or did not adequately represent the nature of particular programs or populations.

Table 5–1 presents an overview of the key advantages and disadvantages of consistent approaches to net savings as identified in the literature, mentioned by interviewees, or experienced in our own work as evaluators. While the project team has characterized these considerations in keeping with the positions of the interviewees, it is our opinion that—other than the lack of a policy driver, referred to earlier—very few of the perceived "disadvantages" and "challenges" are truly obstacles to adopting consistent approaches to net savings in the Northeast Region. We have bolded those that we believe are the most challenging considerations, but even they could be addressed in guidelines that strive for consistency but allow for flexibility. It is also important to note that, as with measurement methods (see Table 3–1), the advantages and disadvantages of consistent approaches to net savings are not of equal weight. For example, the need for air regulators to have regional consistency for purposes of calculating air emissions is extremely important, despite the disadvantages and challenges of doing so.[60] We discuss these advantages, disadvantages, and challenges in more detail in the sections that follow.

[59] As mentioned above, air regulators did not speak with one voice on the issue of adjusted gross vs. net savings.

[60] Before continuing, we must note that, while we specifically asked interviewees separate questions about their opinions on consistency of definitions and consistency of measurement, most chose to give the same response for each concept. Therefore, our discussion below talks about "approaches to net savings."

Table 5–1: Considerations Regarding Consistency in Definitions and Methods for Measuring Net Savings

Perceived Advantages	Perceived Disadvantages	Perceived Challenges
• Know that savings achieved in one program or jurisdiction are comparable to those achieved in another program or jurisdiction • Facilitate evaluation and reporting efforts for regional initiatives or programs that may affect markets beyond a single jurisdiction as well as for program administrators operating in more than one jurisdiction	• Method often depends on unique characteristics of the market, product, program design, or geographic or demographic populations served • **Budget differences may not allow everyone to adopt the same approaches** • **Might stifle creativity in program design, implementation, and evaluation** • Not a high priority compared to improving current methods, sorting out attribution concerns, or rethinking how we assess program success or failure	• **Lack of a policy driver to stimulate development of consistent approaches** • **Getting all the relevant parties to agree on a the details of definitions and methods would be difficult** o Parties would prefer their definition or approach o Would overturn years of regulatory decisions and legislation that underlie current approaches o Would need to meet the needs of everyone • Must be a standard that makes sense, not consistency for its own sake

5.2.1 Perceived Advantages of Consistent Approaches for Net Savings

Our literature review and interviews revealed two primary advantages of implementing consistent approaches to net savings. The first perceived advantage of a standard definition and set of methods is that it would facilitate comparisons of programs both within and across jurisdictions. It would put in place an approach that is simple, consistent across years and across contractors, and easy to implement (*e.g.* Peters and Bliss 2010). Commentators who mentioned this advantage asserted that consistency in the handling of free ridership, spillover, and other adjustments to net savings was especially needed. Overall, consistent definitions and methods would allow, for example, not only easier comparison of net savings between commercial and residential retrofit programs but also between commercial retrofit programs in two or more jurisdictions. As one net savings expert joked, [If we had consistent definitions and methods] "we wouldn't confuse our regulators." Supporters of this opinion believed that, armed with such "apples to apples" comparisons, program administrators and regulators would have a more complete and accurate idea of which programs performed well, which needed revision, and which should be discontinued.

The second perceived advantage of a set of consistent approaches is that it would facilitate evaluation and reporting efforts for regional initiatives such as RGGI, for programs that are expected to have market-level effects that go beyond the boundaries of individual jurisdictions, as well as for program administrators that operate in more than one jurisdiction (for example, First Energy, DelMarVa, Pepco, National Grid, Northeast Utilities, and Unitil, among others, operate in multiple jurisdictions included under the NEEP umbrella). This issue is especially pertinent to air regulators, because they must know that each unit of savings from one

jurisdiction is equivalent to the same unit in another jurisdiction. Speaking to the needs of air regulators, one energy regulator explained, "I think we need to get a uniform system of definitions of reporting practices and of monitoring and verification practices in general so that data can be used across states and can be understood and reliable…. And I think eventually it should be used nationally so that beyond just the Northeast we can share this information." One state cannot adjust by free ridership only while another also accounts for spillover, for example, or the regional calculation of emissions reduction will be deemed unreliable. This could lead to challenges and disputes over the legitimacy of the estimates and reduce the likelihood of securing future program funding from such entities as the EPA and DOE. Furthermore, as one air regulator noted, consistency in monitoring and evaluation of different programs within a jurisdiction and across the region is important for understanding how the needs of the air community fit with those of the energy community. Without this consistency, the two communities may not be able to coordinate their efforts in the manner needed to achieve ambitious reductions in greenhouse gas emissions.

While most interviewees who support consistent approaches did not volunteer what the consistency would entail, one energy regulator suggested that "The definitions and methods should be standardized, and the standard should meet our [*i.e.* energy regulators'] needs. Questionnaires, comparison studies, and economic models might be among the things to be standardized." Another energy regulator expressed, "I think there should be a regional definition of what is a free rider, what is spillover, and regional guidelines of how the information will be reported. I guess the most comprehensive way is to have gross savings, free ridership, spillover, and net all reported separately, so we would have every piece."

Yet for each of these characteristics that some interviewees described as advantageous and desirable, other interviewees cited as problematic and disadvantageous.

5.2.2 Perceived Disadvantages of Consistent Approaches for Net Savings

The literature review and interviews revealed three perceived disadvantages to instituting consistent approaches to net savings across the region. Throughout this discussion, it is important to note that we did not inquire about the *degree* of consistency. Interviewees largely responded with the assumption that there would be *one definition* and *one method* that would apply to all programs and jurisdictions. Furthermore, they assumed the continued focus on program evaluation as opposed to overall portfolio evaluation. If we relaxed the assumption and allowed for diversity within a set of consistent *guidelines* and for portfolio level evaluations, then some of the perceived disadvantages may become moot. For others, the design of the consistent framework could explicitly take the considerations into account, lessening their potentially problematic nature. We discuss such mitigations in greater detail below.

The first perceived disadvantage is the belief among many commentators that the method of measurement should depend on the unique characteristics of the product, program design, or populations served. As one program administrator explained, "Depending on what you are

evaluating, factors come into play that may differ by program and technology. Getting a method that covers every case would be difficult." Another program administrator staff evaluator stated, "Not everyone can do the same thing. Some have more evaluation money than others and can look more deeply into net savings." Other interviewees pointed out that something as simple as the different ways in which customers shop in New York City versus upstate can force evaluators to turn to different approaches to measuring net savings. Low-income programs serve as another example. Many low-income weatherization programs do not calculate free ridership because it is unlikely that the respondents would have adopted the measures on their own, due to the lack of disposable income to do so. In such cases, identifying what would likely be a very low percentage of free riders would waste limited resources and not be a cost-effective use of program funds.

The second perceived disadvantage raised in the literature and by interviewees is that instituting consistent methods could stifle creativity of program design, implementation, and evaluation.[61] Consider upstream CFL programs. These programs pay incentives to manufacturers and retailers, who then reduce prices on store shelves. The programmatic shift to this approach throughout most of New England in the mid-2000s led to huge increases in the number of CFLs being sold in program areas. The structure of upstream programs, however, seriously limited the ability of evaluators to estimate free ridership for a variety of reasons, but the key ones are that participants do not know they were taking part in a program and that free ridership is built into the design in that a CFL purchaser has little choice but to buy a discounted product when that is the model carried by the store. In response, evaluators turned to market-based approaches that did not allow for the separate estimation of free ridership and spillover. Had program administrators been penalized for not providing these free ridership estimates, this very effective CFL program design may not have been implemented by the first states to adopt the approach. Now, many states and jurisdictions have copied the design because they recognize its effectiveness, yet some of them have not revised their reporting rules about free ridership for CFL programs. Program administrators and evaluators are struggling to find methods that will yield reliable estimates of free ridership in the face of a program design that is not amenable to such measurement. The situation is contributing to nationwide confusion and concern about measurement of NTG ratios for CFLs in a rapidly changing market.

Finally, a number of interviewees argued that developing consistent approaches to net savings should remain low on the priority list of the energy efficiency community, in large part because current methods do not capture the full range of program impacts and will not serve the field well as programs continue to evolve. As one net savings expert stated, "It's more important to develop a new paradigm [for net impact evaluation] for the future than to consolidate the old, antiquated paradigm." A program administrator theoretically supported the idea of a consistent approach to net savings, but held that we first have to develop more reliable estimation methods

[61] In addition to multiple interviewees, Scriven makes this point in his debate with Cook regarding funding agencies that mandate the use of randomized control experiments to assess education programs (Cook *et al.* 2010).

to avoid promoting flawed methods that do not adequately capture the entire range of program effects. A different net savings expert echoed these perspectives, "We first need to identify and classify various types of effects and figure out how to measure them before coming up with [a] consistent [approach to] net savings. We need to spend more time thinking about the nature and size of effects in order to come up with a better understanding of what should be included in net savings and how to measure it."

5.3 Perceived Challenges to Implementing Consistent Approaches for Net Savings

Whether or not interviewees supported consistent approaches to net savings, most of them felt that implementing a consistent approach would face two potential challenges.

The first perceived challenge may be summarized as follows: Even if the various parties and jurisdictions could agree in theory to the concept of consistent approaches to net savings, it would be difficult to get them all to agree on the nature of the consistent approach to be used in practice. In fact, some commentators argued, this is why previous calls for consistency have gone unheeded. It is tempting to dismiss this perceived challenge as a self-interested aversion to change or even the energy efficiency community's own version of a "turf war." As one program administrator quipped, "It would be nice if everyone agreed with me and my definition." Despite the frequent jokes of this sort, the interviewees more seriously explained that the tendency to want "my approach" is less about "turf" or an aversion to change and more about the probable hurdles of implementing the changes. In particular, the diversity of net savings approaches often reflects cumulative regulatory and legal decisions over many years, and these decisions have led to divergent practices throughout the region. Consistent approaches, then, may realistically have to come from interstate agreements, as one energy regulator explains: "Right now it's very easy for each state to follow its own practices and ignore all the others. But, any kind of interstate agreement will change that." Furthermore, as mentioned above under perceived disadvantages, the diverse nature of the programs, products, and customers involved would make it difficult to find a single, consistent method that meets everyone's needs, although a consistent framework that allows for flexibility in methods would alleviate this concern.

The second challenge is to ensure that the consistency has a real purpose and meaning behind it and that the standards make sense; consistency cannot be pursued for its own sake. Given the increasingly regional approach to reducing energy use and demand as well as greenhouse gas emissions, it appears that there may now be a driving need for consistency.

5.4 Overcoming Barriers to Consistency in Approaches for Net Savings Definitions and Methods

The interviewees raised important considerations that must be addressed in any attempt to develop and promote a consistent framework for net savings in the Northeast. However, as

mentioned above, the responses to our interview questions generally assumed that a consistent approach meant "one size fits all," rather than a consistent yet flexible framework. In this section, we address the perceived concerns raised by the interviewees and provide an initial response of how they may or may not affect the development of a consistent approach for defining and measuring net savings in the Northeast Region. In this discussion we focus on free ridership and spillover, the two adjustments to net savings that sparked the greatest concern in the literature and among interviewees.

5.4.1 Consistent Definitions

Interviewees more immediately embraced the idea of a consistent definition of net savings than the idea of consistent methods. Most believed it would be a good idea for jurisdictions in the Northeast to define net savings in a similar manner, without the current variations in inclusion or exclusion of free ridership and spillover (or sub-types of each of these). The main consideration raised about consistent definitions was simply getting the various jurisdictions to move from their current definitions to a single regional definition that would require some jurisdictions to change their definitions, at least to an extent.

Based on our research findings, the project team believes that if the region decides that a consistent definition of net savings is a good idea, the Forum could lead the process of developing this definition. Our research suggests that many program administrators, energy regulators, and air regulators are open to the idea, which is a strong first step to making consistent definitions a reality.

5.4.2 Consistent Methods

While our research suggested that the energy efficiency community at least theoretically embraces the idea of a consistent regional definition of net savings, the literature identified and interviewees voiced a larger number of issues about consistent methods. As one energy regulator stated, "I'm not sure I'm at the point where I think the methodology should be dictated, but it would be useful to have similar approaches for similar programs in different jurisdictions." It is the opinion of the project team that only a few of these considerations serve as actual obstacles to the development of consistent methodological guidelines for net savings, and even these could be mitigated through a carefully planned and implemented framework involving regional cooperation. In this section, we address each of the considerations raised and how they might be addressed in a consistent net savings framework. The quotations below are drawn from topics addressed in Sections 5.2.2 and 5.3; they sometimes summarize the issues discussed above and other times represent direct quotations from interviewees.

"The method of measurement should depend on the unique characteristics of the product, program design, or populations served." This consideration follows from the assumption that consistency means one size fits all. Consistency put into practice as flexible guidelines, in contrast, reduces the potential challenge of this concern by providing methodological options for different contexts. The case remains that because options are based on generalities, not specifics,

two organizations might opt for two different options for programs of the same type, but this tendency could be lessened through education on how to use the guidelines.

"Not everyone can do the same thing. Some have more evaluation money than others and can look more deeply into net savings." The objection is not to having a minimum requirement for a consistent approach, but rather to imposing a standard that is too high for everyone to reach. The project team believes this is a critical, but not insurmountable, challenge to consistency. Many of the more valid methods are often—although not always—the more expensive methods. A situation could arise in which a jurisdiction could always only afford less valid options. One possible solution would be for program administrators among jurisdictions to combine resources in joint evaluations, a practice already common among in the Northeast. Joint evaluations can be designed in a manner that would allow for separate estimates of net savings for each jurisdiction if needed.

"[Some programs such as low-income ones are expected to have low free ridership or spillover]. In such cases, identifying [free ridership or spillover] would waste limited resources and not be a cost-effective use of program funds." The project team believes that this consideration can be addressed through decision rules within guidelines for consistency that are sensitive to the characteristics of programs such as low-income ones.

"Instituting consistent methods could stifle creativity of program design, implementation, and evaluation." The consideration raised here is not allowing the consistent methods to drive program design, implementation and evaluation. The project team concurs with this concern but also asserts that evaluation—including net savings evaluation—must be incorporated into program design and implementation; evaluation cannot be treated as an afterthought. Guidelines for consistent net savings estimation would need to be flexible enough to allow innovation and creativity in program design, implementation, and evaluation while still ensuring that net savings evaluation is not forgotten in the process. Pilot projects, furthermore, offer an opportunity to test innovative approaches, including not only in program design and implementation but also in evaluation. While savings estimated for some pilot projects may not achieve the rigor needed for reporting regional energy use or greenhouse gas reductions, when successful pilots are scaled up, net savings evaluation that adheres to the framework could be incorporated into the program design and implementation.

"Had program administrators been required to calculate free ridership and spillover separately, they may not have adopted this very effective CFL program design [upstream programs]" This is an objection to having to report free ridership and spillover as separate estimates of net savings, not to net savings estimates *per se*. A consistent but flexible set of guidelines could allow for the reporting of net savings in a manner that does not require the isolation of free ridership or spillover when doing so is not appropriate to the program design or implementation strategies.]

"It's more important to develop a new paradigm [for net impact evaluation] for the future than to consolidate the old, antiquated paradigm." This also is not an objection to net savings

estimates, but to relying on current methods in the development of consistent guidelines to measuring net savings. While the industry may soon find that it needs vastly different approaches to program design, implementation, and evaluation to achieve ambitious reductions in energy use and greenhouse gas emissions, such a shift will not happen overnight. Moreover, one cannot develop new evaluation approaches for program designs that do not yet exist. Therefore, in the near future, guidelines for consistent methods for net savings could be developed that draw upon best practices within current approaches while still allowing for the development of innovative techniques that will meet possible future shifts in the paradigm of energy efficiency programs.

6 Next Steps, Recommendations, and Future Research

The research presented in this scoping paper demonstrates that the definition, measurement, and uses of net savings have long challenged the energy efficiency community and continue to do so today. Some evidence suggests that these challenges have recently been exacerbated by an increase in regional cooperation on energy efficiency and climate change programs and policies, an expansion of the audiences for energy savings estimates, and a proliferation of influences (*e.g.*, programs, media coverage, and economic policies) affecting energy efficient action. Together these persistent and growing complications lead to confusion and disagreement over such issues as the best way to demonstrate that a program *caused* savings to occur, whether the air regulation community will need adjusted gross or net energy savings to estimate emissions reductions, and whether the Northeast should pursue consistent definitions and measurement approaches to net savings.

Despite the continued confusion and debates, our research does support the continued use of net savings estimates for certain specific purposes, as follows:

- *Assessing the degree to which programs bring about a reduction in energy usage and demand*. While this is a controversial use of net savings estimates, the research supports the continued use of net savings as *one* of numerous measures that should be given serious consideration in the assessment of program success, at least until a suitable alternative is developed beyond gross or adjusted gross savings that recognizes that some energy and demand savings would have happened without the program.
- *Uncovering fraudulent program implementation practices*. Cases have been documented in which program implementers have claimed savings from activities that they clearly did not influence, including installations that occurred prior to any interaction with the program and random downward fluctuations in energy use in excess of what accrued to any program activity
- *Gaining insight into how the market is changing and transforming over time* by tracking net savings across program years and determining the extent to which free ridership and spillover rates and net-to-gross ratios have changed over the period.
- *Understanding better how the market responds to the program* and using the information to inform modifications to program design, including measure eligibility and targeted marketing. Later, these program modifications would again be subject to net savings evaluation, in an adaptive management process.

6.1 Recommendations and Research Needs

In addition to these four uses, the research also points to a series of recommendations and research needs related to pursuing consistent approaches to defining and measuring energy savings—net and gross—throughout the Northeast, securing higher quality data, determining program attribution, and aligning the research needs and activities of the energy efficiency

community with those of the air regulation community. We present each recommendation below, and expand on them in the main body of the report.

Recommendation 1: The Forum should lead the process of developing a consistent definition of *adjusted gross savings* in the Northeast Region. The research activities conducted for this scoping paper revealed that different jurisdictions in the Northeast Region and the nation vary in the degree to which they measure not only free ridership and spillover but also such parameters as rebound, leakage, hours of use, installation rates, and persistence rates, among others, and whether these additional parameters are considered to be adjustments to gross savings or elements of net savings. It is the *opinion* of the project team that all of these concepts *except free ridership, spillover, and rebound* constitute adjustments to gross savings as they measure what actually *has* happened, as opposed to free ridership, spillover, and rebound which involve the measurement of what *would have* happened in the absence of the program. Ultimately, the Forum will have to decide if it agrees with our recommendation and, if so, which of the possible adjustments will be included in its definition of adjusted gross savings.

Related Research Needs for Recommendation 1:

- Identifying the adjustments that are currently submitted to the ISOs for reporting in support of the forward capacity markets.
- Examining the components that factor into realization rates, which also are not consistently defined from program to program, jurisdiction to jurisdiction, or evaluation to evaluation.
- Evaluating the degree to which our recommended components for adjusted gross savings meet the needs and requirements of the energy efficiency community in the Northeast, given that our efforts focused on net savings and not adjusted gross savings.

Recommendation 2: After developing a consistent definition of adjusted gross savings, the Forum should develop a definition of *net savings* for the Northeast Region. This definition would be in relation to—but separate from—that of adjusted gross savings. The Forum would need to consider which concepts should be included in this definition, particularly focusing on the treatment of free ridership and spillover.

Related Research Needs for Recommendation 2:

- Enumerating in detail the similarities and differences between current definitions of net savings in use throughout the Northeast.
- Understanding the reasons for differences in definitions of net savings and potential ways to bring the definitions into alignment while meeting the needs of the diverse energy efficiency community in the Region.

Recommendation 3: The Forum and its members and allies should consider advocating for legal requirements for manufacturers, retailers, and distributors to provide national sales and shipment data for key equipment and products—reported by size and efficiency, ideally at the county, or at least state, level. Legally requiring reporting of these data may

provide the only avenue for accurate estimation of net savings from upstream market transformation programs—including capturing the savings resulting from cumulative program activity from prior years. Higher-quality data will not only improve estimation from quantitative approaches, but they could also provide more dependable information for use in qualitative techniques such as structured judgment (Delphi panels), weight-of-evidence, and historical tracing approaches. Note that this recommendation does not require further research per se, but, instead would focus on advocacy perhaps accompanied by a research project to facilitate the advocacy and develop supporting materials on the part of the Forum and its members.

Recommendation 4: Program administrators should keep records of program activity by year (*e.g.*, budgets, measure covered, participants served, previous estimates of net savings, *etc*.), including in any possible comparison areas. Such data will assist in the development of net savings estimates by allowing for the estimation of cumulative and spillover effects over time, thereby providing a more complete assessment of program impacts. Although this recommendation must largely be implemented by program administrators, the Forum could assist program administrators in the development of data collection and storage tools, perhaps even developing a database that contains such information for easy reference and comparison across areas and years. Currently, the Consortium for Energy Efficiency keeps track of program data for members who respond to their annual request for information, and their experiences should certainly be taken into account. Note also that, while this recommendation is not dependent upon the adoption of Recommendations 1 and 2, consistency in definitions of adjusted gross and net savings would facilitate the keeping of such records and the development of a database.

Recommendation 5: The Forum should also clarify the definition of *attribution* and the degree to which programs must differentiate the impact of their activities from the impacts of other programs and efforts designed to bring about the same or similar actions or outcomes, and from contextual factors such as energy prices and the economy. The term "attribution" is consistently applied to capture the impacts *caused* by program activity, but is increasingly being used to refer to separating the effects of *this* program from the effects of all the other programs or pro-efficiency messages to which participants have been exposed, and from the effects of contextual factors such as energy prices and the economy, during a specific period of time. While it may be important to assess the impacts brought about by the program, the energy efficiency community should be careful not to lose sight of the goal of reducing energy use by focusing too narrowly on which program gets the credit, particularly as various programs and influences may have synergistic effects such that the whole of their combined impact is more than the sum of their individual impacts.

Research Needs for Recommendation 5:

- Delineating the percentage of participants that are "repeat players"—that is, those taking part in the same program in different funding cycles or taking part in multiple programs offered by the same program administrator in the same or different funding cycles.

- Comparing the total savings of repeat players to the aggregation of savings among comparable individual program participants (*i.e.,* the savings from Participant A from program X plus the savings from Participant B from Program Y, and so forth).
- Identifying the variety of potential influences outside of program activity that may affect program participation and assessing whether direct causal links can be established between those other potential influences and energy saving actions.

Recommendation 6: The Forum should encourage the energy efficiency community—particularly energy regulators—to expand its assessment of program success from a focus on net savings, particularly free ridership, to the inclusion of additional factors that may more accurately capture the full range of program impacts, including non-energy impacts such as jobs, improved health, and increased productivity. Net savings is a valuable measure of the amount of savings the program has achieved that would not have happened otherwise. But, as this scoping paper has documented, net savings measures are not infallible, nor do they fully capture the wide range of impacts that may result from energy efficiency program activity. Therefore, while we recommend the continued consideration of net savings estimates, we also suggest that additional impacts be taken into account when assessing programs and determining any reward or penalties to be paid.

Research Needs for Recommendation 6:

- Exploring approaches that best capture the cumulative—that is, long-term—effects of market transformation programs on the targeted markets.
- Identifying the impacts beyond energy savings that are most crucial to determining program performance and ensuring that ratepayer and taxpayer funds are being used responsibly
- Developing a prioritization scheme that may assist energy regulators in expanding their assessment of program performance beyond net energy savings and its components

Recommendation 7: The Forum should decide if it supports the development of consistent methodological approaches to estimating net savings for the Northeast, and, if so, take the actions necessary to develop regional guidelines for consistent methods. This recommendation involves three different steps. First, the Forum should discuss the reasons for and against developing consistent approaches to net savings estimation and decide whether consistency should be pursued in the Northeast. Second, should the Forum recommend in favor of consistency in methods for estimating net savings, it should begin working toward consistency by first developing a framework for how to achieve this consistency. The framework would set forth initial suggestions on how program administrators could begin moving toward consistency. The framework could also identify some of the key considerations that program administrators and evaluators should take into account when choosing from among the various methods for measuring net savings. This framework would serve as an interim step toward consistency while the Forum waits for the results of research projects that would be needed before embarking on the third and final step: the development of guidelines for consistent methodological approaches

to net savings. The guidelines would provide more explicit recommendations concerning which methodological approaches to use in specific situations. The guidelines should also allow for the introduction of new methods, with a process or criteria for establishing their reliability, validity, and rigor. Although flexible, the guidelines should avoid the trap of "anything goes" by specifically identifying the best approaches to be pursued given varying levels of resources and by identifying those approaches that should be avoided except in very limited circumstances.

Research Needs for Recommendation 7:

- Understanding the range of methods available for estimating net savings, their strengths and weaknesses, and their ability to provide the types of estimates needed by the energy efficiency community (and perhaps the air regulation community) to ensure that any guidelines for consistent methodological approaches to net savings focus on the best net impact evaluation practices available.
- Exploring the existing academic and energy efficiency research on the psychological and sociological processes that influence estimates of net energy savings in an effort to understand more fully the ways in which these processes may affect how participants respond to self-report questions about their past actions and likely behavior in the absence of the program.
- Examining the potential of macroeconomic approaches for estimating the impact of program activity on net energy savings. Such approaches are new, and it remains to be seen if they will prove to be among the new "best practices" in net savings evaluation.

Recommendation 8: The Forum should facilitate the development of a working group comprising members of the energy efficiency community, the system planning community, and the air regulation community with the ultimate goal of developing approaches to measuring energy savings and resultant reductions in greenhouse gas emissions in a manner that is mutually acceptable to and feasible for all three communities. This recommendation rests on the assumption that the federal government will likely adopt guidelines on how to account for emissions reductions resulting from energy savings achieved through efficiency programs. It further assumes that the energy efficiency community would prefer to be an active part of the development of these guidelines rather than responding to guidelines set forth by the air regulation community. Given these assumptions, such a working group would have three tasks. The first task would be to develop a mutual understanding of the terminology used in each field and to identify compatible concepts across both fields. Where concepts are not compatible, the working group would seek to bring them into closer alignment to the extent possible while still meeting the needs of the two communities. The second task would be to review methodologies for estimating baseline emissions reductions, gross energy savings, and net energy savings, making certain that the baselines for emissions reductions do not include projections of energy savings expected to come from efficiency programs and that the methodologies for estimating energy savings fulfill the SIP verification processes. In the final task, the working group would deliver a set of recommendations to the EPA, and state energy

and air regulation agencies on how best to account for energy savings from emissions reductions in a manner that is mutually agreeable to both the energy efficiency and air regulation communities.

Research Needs for Recommendation 8:

- Examining the possible methods or approaches for translating energy efficiency impacts into measureable pollution reductions from power plants by time of day and time of year.
- Exploring possible approaches to measuring *net energy savings* that would meet the requirements for reliability and precision set forth in SIPs. This may involve the use of technology such as advanced metering (if a method to translate the information to net savings were developed) or the increased use of randomized controlled experiments in estimating net savings.

The project team believes that the energy efficiency community—and probably the air regulation community as well—should acknowledge the fact that energy efficiency and emissions reductions programs are not wholly responsible for all the savings and emissions reductions that may be achieved through the adoption of the devices or behaviors promoted by the programs. The estimation of net impacts—be they savings, emissions reductions, jobs created, water saved, and so on—is a way of making this acknowledgement. Currently, the measurement of net impacts occurs at the measure or program level, with the results sometimes being aggregated to a group of programs or to the entire portfolio offered by a program administrator. However, we recognize that demonstrating attribution has become increasingly challenging in light of numerous programs and factors that may influence the targeted behaviors Moreover, jurisdictions at the municipal to state levels—and perhaps one day at the federal level—have set ambitious goals to reduce energy use and greenhouse gas emissions. Given the uncertainty inherent in measuring the counterfactual of net savings, some members of the energy efficiency community are beginning to ask if we need a radical rethinking of how we go about estimating net impacts in general and net savings in particular. It is beyond the scope of this paper to describe what such a radical rethinking of impact evaluation methods might entail, but we do offer a few interim recommended alternatives to *measuring* net savings that nevertheless still recognize that program activity does not necessarily yield all the measured gross savings. This brings us to the final recommendation.

Recommendation 9: The Forum should consider the potential of using a deemed or negotiated net savings approach for crediting energy savings—or emissions reductions—to a program or portfolio. In such an approach, program administrators and regulators would draw on available evidence and additional indicators of program impacts to help decide on the percentage of gross savings that can be claimed by a program or portfolio after it has carefully demonstrated that the program activities have strongly contributed to the desired outcomes. The credibility of the negotiated net savings figure would depend on the type, amount, and quality of the information informing it; such information could include program tracking data, adjusted gross savings estimates, net savings estimates derived from periodic research (possibly multiple

approaches), the sales/shipment data mentioned in Recommendation 3, market research to assess the state of the market, energy intensity by sector over time, and more. The percentage could be developed through a Delphi panel or using other structured expert judgment approaches. Importantly, the approach itself, and not just the number, should be negotiated and agreed on beforehand; it is even possible—as with the recent decision in Arizona to credit up to one-third of the savings from utility C&S efforts toward its 2020 Energy Efficiency Standards (EES) target, provided they make a credible effort—to decide on the number beforehand. The conclusion would still be subjective, and its precision—how close it is to the real value of net savings—would be uncertain because achievement of the goal could be questioned on the basis of the counterfactual as well as factors such as economic recessions, reduction in certain types of manufacturing, changes in energy prices, and so forth. Even so, the negotiated agreement approach avoids depending entirely on controversial measurements of the counterfactual of net savings.

To conclude, the definition and measurement of net energy savings remains a controversial issue in the energy efficiency community. The controversy surrounding the issue presents challenges to the task of achieving consistency in definitions and measurement of net savings throughout the Northeast. Some commentators do not believe consistency is needed or desirable, while others perceive the appeal of consistency but do not believe it can be achieved in practice. Other commentators, however, support the idea of increased regional consistency in definitions and measurement of net savings, citing increased cooperation in reducing energy use and greenhouse gas emissions across state lines and the fact that many program administrators operate in multiple jurisdictions. This paper has sought to clarify some of the issues surrounding net savings definitions and measurement and the pursuit of regional consistency. The recommended action and research needs follow directly from the observations identified in the literature, by interviewees, and by various Forum and Steering Committee members. It is our hope that the findings and recommendations will help the Forum—and the regional energy efficiency community—come to some agreement regarding the definition and measurement of adjusted gross savings and net savings, whether to pursue consistency in the near future, and the degree to which the needs of the air regulation community may be a driver of both.

Appendix A Interview Guides

The two interview guides are listed on the following pages.

A.1 Interview Guide for Energy Efficiency Experts

My name is [INSERT] and I'm calling from [INSERT] on behalf of the Northeast Energy Efficiency Partnership's Evaluation, Monitoring, and Verification (EM&V) Forum. The Forum is exploring the issue of "net energy savings" and various issues surrounding it, and they believed that you may have some unique insights to share related to this topic. Your responses will be kept confidential, and we will make every effort to hide your identity when reporting the results. [SET UP OR CONTINUE WITH INTERVIEW] Do I have your permission to tape the interview? This is for transcription purchases only. [Y/N]

We are interviewing you because you are considered an expert in your field. While a few of our questions explicitly ask about the perspective of your organization, we generally prefer that you respond from your own personal perspective unless directly asked to speak for the organization. If you decide to provide the opinion of your organization, please explicitly note this in your response.

1. How does your **organization** define "net savings"? [PROBE: Do you include free riders (if so which kinds), spillover (if so, which kinds?), net-to-gross ratios, savings exclusively attributable to the organization's program, other (specify and explain)? Do you **personally** define net savings in the same way? If not, how do you personally define net savings and why do you adhere to this alternative definition?

2. What do you think is the appropriate role of net energy savings estimates in energy efficiency evaluation? [PROBE: compared to gross energy savings estimates; importance, value, need for net–savings [if more specificity needed could prompt for attribution, causation, program effectiveness, appropriate use of funds, etc.]

3. Which aspects of measuring net energy savings are most well developed, and which are most in need of improvement? [PROBE: definitions, measurement, use of results overall, related to individual components such as free ridership, participant spillover, and/or non-participant spillover, or related to particular program types.]

4. [USE THE FOLLOWING GRID AS GUIDE TO ASKING NEXT QUESTIONS]

Audience Group	Who are the audiences for net savings estimates? [PROBE: the following if not volunteered. Indicate if respondent thinks group is an audience.] probe about consistency—state, regional/more than one state, national	[FOR EACH GROUP RESPONDENT IDENTIFIES AS AUDIENCE, ASK] How might [INSERT GROUP] use net savings estimates to influence or redirect program design or implementation? [That is, what implications do their responses to the estimates have for program design, implementation, revision?]
Program planners	Y/N	ASK
Program implementers	Y/N	ASK
Energy regulators	Y/N	ASK
Air regulators	Y/N	ASK
Legislators and other policy makers	Y/N	ASK
Stakeholder groups	[IF YES, ask to name which types of stakeholder groups; If necessary, probe as examples: environmentalists, rate payer alliances, social justice advocates, industry groups]	DO NOT ASK
Ratepayers	Y/N	DO NOT ASK
General public	Y/N	DO NOT ASK

5. In what way do the current approaches to estimating net savings meet current reporting and policy needs? What about evolving policy needs? [PROBE: specifics on what is missing, if anything]? [USE THE FOLLOWING GRID AS CHECKLIST TO MAKE SURE HIT ALL CATEGORIES]

Policy or Reporting Need	**Current Needs**	**Evolving Needs**
Energy savings goals (programmatic)		
Energy savings goals (state or regional)		
Air emissions		
Carbon emissions		
Other GHC emissions		
Other reporting needs		
(specify other need)		
(specify other need)		

6. What level of resources should go into estimating net savings, compared to other evaluation, monitoring, and verification activities? [PROBE: estimating gross savings; assessing non-energy benefits; market characterization and assessment; process evaluation, variations by program type or structure, time evaluators spend estimating net savings versus other goals of evaluations] Why? How much should cost weigh into deciding among various approaches to estimating net savings?

7. In the changing environment in which some states call for decreased energy use rather than a decreased rate of growth—are net savings estimates important? Do net savings estimates have a role to play? If so what? If not, why not? Is it vital to determine causation and to attribute particular savings to individual programs in this environment? Why or why not? [For example, Massachusetts has goals of reducing electricity use by 2.4% annually and natural gas use by 1.15% annually for the next three years. Climate change goals entail reductions in carbon output, not just reduced growth.]

8. Are there any situations in which net savings estimates are not appropriate? If so, what situations, and why are net savings estimates not appropriate?

9. What would be the benefits and drawbacks of moving toward a consistent *definition* of net savings and its various components? [For example, consistent inclusion or exclusion of various types of free riders and spillover, market effects, relation to codes and standards;. regional, state, national, ISO level, air quality zones] Should the Northeast region attempt to have regional guidelines/recommendations for *defining* net savings? Why or why not? When should that happen? What would you recommend as the best regional definition of net savings?

10. What would be the benefits and drawbacks of moving toward consistent *methods* across states for estimating net savings? Should the Northeast region attempt to have regional guidelines/recommendations for *calculating* and *reporting* net savings? Why or why not? [IF THERE SHOULD BE GUIDELINES] When should that happen? At what level is consistency needed? [PROBE: regional, state, national, ISO level, air quality zones] What would such consistency look like? [PROBE: Is it one size fits all across states and programs, a set of parameters to be measured but exactly how varies across programs or states, a mixture? Explain and provide examples.]

Thank you for your time today. The report will be posted on the NEEP EM&V website when the project is finalized.

A.2 Interview Guide for Air Regulators

Energy efficiency program administrators and evaluators are frequently called on to estimate the impact the program had above and beyond what would have happened in the absence of the program. This is also called the net impacts of the program. Net impacts can be estimated using many different methodologies and are the source of some controversy.

1. As an air regulator, are you interested in current values of air emissions (how much is emitted) or reductions in air emissions, which necessarily compare to some baseline? Both? Something else entirely?
    a. [If baseline] What baselines do you use?
    b. Are you aware of any differences by region regarding interest in current value vs. reductions?
2. Is it important to understand what led to observed reductions in air emissions or estimated greenhouse gas reductions, or is it important to document only that the reductions have occurred? Is it important for the air quality community to know the source of GHG reductions—say whether reductions came from utility energy efficiency programs or from other emission reduction initiatives?
3. One air regulator we spoke with has told us, in effect: "What we need from the energy efficiency community is an estimate of MWh (or therm) savings across the region that has been consistently measured by all parties reporting savings in the region. We will take that estimate and convert it to emissions and do our work. We'll yield to the energy efficiency community issues of how the estimates are calculated." Does this statement reflect your views? The views of your colleagues? Why or why not?
4. How do you need energy savings reported? Are simple estimates of MWh and therm savings sufficient, or do you need savings reported by various categories, under certain conditions, or by when the savings occur (time of the year and time of day)? [IF THEY DON'T SEE THE NEED FOR SAVINGS BY TIME OF DAY AND YEAR, SAY: "One reason for having savings estimates by time of year and time of day is that, for example, savings from residential lighting programs might occur mostly on winter evenings and mornings when much of the demand is met by generation sources with low or no emissions, such as nuclear or hydro, whereas savings from air conditioning programs might occur mostly on hot summer days when much of the demand is met by peaking plants with higher emissions."]
    a. Are there air resource terms or reporting formats that the energy efficiency community should become conversant in? If so, what?

Appendix B: Bibliography of Works Cited or Reviewed

Adams, Alyce, Stephen Soumeria, Joanathan Lomas and Dennis Ross-Degnan. 1999. "Evidence of Self-report Bias in Assessing Adherence to Guidelines." International Journal for Quality in Health Care 11(3):187-192.

Agnew, Ken, Paula Ham-Su, and Miriam Goldberg. 2006. *Focus on Energy Statewide Evaluation: Business Programs: Lighting and Motor/Drive Channel Market Effects Contract Metrics Assessment.* Prepared for State of Wisconsin Department of Administration Division of Energy. Madison, WI: KEMA.

Allcott, Hunt, and Sendhil Mullainathan. 2010. "Behavior and Energy Policy." *Science* 327: 5970.

Allcott, Hunt. 2009. *Rethinking Real Time Electricity Pricing*. MIT Center for Energy and Environmental Policy Working Paper.

Barnes, Harley. 2004. "Taking Some of the Uncertainty out of Retailer Self-Reported ENERGY STAR Market Share." Pensacola, Florida: *Proceedings of the Annual Meeting of Association of Energy Service Professionals*, December.

Bliss, Ryan, Jane Peters, Ben Bronfman, Matei Perussi, Rick Ogle, and Ross Notebaart. 2009. *Impact and Process Evaluation of the 2006-2007 Building Efficiency Program*. Prepared for the Energy Trust of Oregon. Portland, OR: Research Into Action.

Blumstein, Carl. 2009. *Program evaluation and incentives for administrators of energy efficiency programs: Can evaluation solve the principal/agent problem?* University of California Energy Institute.

Cadmus Group, Inc., KEMA, Itron, NMR Group, Inc., and A Goett Consulting. 2010. *Compact Fluorescent Lamps Market Effects Final Report.* Prepared for the California Public Utilities Commission. April.

California Public Utilities Commission Energy Division. 2007. *Guidelines for Estimating Net-to-Gross Ratios Using the Self-Report Approaches.*

Connell, James P. and Anne C. Kubisch. 1998. *Applying a Theory of Change Approach to the Evaluation of Comprehensive Community Initiatives: Progress, Prospects, and Problems.* The Aspen Institute.

Cook, Gay. 2008. "Attribution Methodology Wars: Self-Report Methods versus Statistical Number Crunching—Which should win?" Pacific Grove, CA: *Proceedings of the 2008 ACEEE Summer Study on Energy Efficiency in Buildings*, August.

Cook, Thomas D., Michael Scriven, Chris L. S. Coryn and Stephanie D. H. Evergreen. 2010. "Contemporary Thinking about Causation in Evaluation: A Dialogue with Tom Cook and Michael Scriven." *American Journal of Evaluation* 31(1):105-117.

Degens, P., S. Castor, M., and McRae, R. 2009. "Influence and Intention as Determinants of Free Ridership." Poster at *International Energy Program Evaluation Conference*, August.

DeGroff, Amy, Michael Schooley, Thomas Chapel, and Theodore H. Poister. 2010. "Challenges and Strategies in Applying Performance Measurement to Federal Public Health Problems." *Evaluation and Program Planning* 33:365-372.

Department of Energy. 2010. *SEP Program Notice 10-06: DOE Recovery Act Reporting Requirements for the State Energy Program (SEP) Effective March 1, 2010.*

Donaldson, Stewart I. and Elisa J. Grant-Vallone. 2002. "Understanding Self-report Bias in Organizational Behavior Research." *Journal of Business and Psychology* 17(2):245-260.

Drew, Tim. 2010. "Net Savings for California's Investor Owned Utility Programs." Presented at *US EPA Webinar—Understanding Net Savings*, April.

Dyson, Christopher and Miriam Goldberg. 2007. *The Gift That Keeps Giving: A Structured Approach for Measuring Participant Spillover*. Chicago. *Proceedings of the International Energy Program Evaluation Conference*, August.

ELCON. 2008. *Utility Energy Efficiency Programs: Too Cheap to Meter? A Policy Brief of the Electricity Consumers Resource Council.*

Erickson, Jeff and Mary Klos. 2009. "Free Ridership: Arbitrary Algorithms vs. Consistent Calculations." Summit Blue Consulting. Portland, OR: *Energy Program Evaluation Conference.*

Eto, J., R. Prahl, and J. Schlegel. 1996. *A Scoping Study on Energy-Efficiency Market Transformation by California Utility DSM Programs.* Berkeley, CA: Lawrence Berkeley National Laboratory.

Feldman, Shel, Lisa Wilson-Wright, Lynn Hoefgen, and Angela Li. 2005. "Modeling the Effects of U.S. Energy Star Appliance Programs." *ACEEE Summer Study: What Works and Who Delivers?*

Ferraro, Paul J. 2009. "Counterfactual Thinking and Impact Evaluation in Environmental Policy." In Birnbaum, Matthew and Per Mickwitz (eds.) *Environmental Program and Policy Evaluation: Addressing Methodological Challenges. New Directions for Evaluation* 122: 75-84.

Fitz-Gibbon, C.T. and L.L. Morris. 1987. *How to Design a Program Evaluation (CSE Program Evaluation Kit).* Newbury Park, CA: Sage.

Foote, Susan Bartlett, Beth A. Virnig, Robert J. Town, and Lacey Hartman. 2008. "The Impact of Medicare Coverage Policies on Health Care Utilization." *Health Services Research* 43(3).

Friedmann, Rafael. 2007. "Maximizing Societal Uptake of Energy Efficiency in the New Millennium: Time for Net-to-Gross to Get Out of the Way?" Chicago: *Energy Program Evaluation Conference.*

Glacier Consulting Group, LLC and Ralph Prahl Associates. 2007. *Comprehensive CFL Market Effects Study, Final Report.* Prepared for State of Wisconsin Department of Administration Division of Energy.

Goldberg, Miriam and G. Kennedy Agnew. 2003. "Dynamic Modeling of Market Effects and Spillover with Limited Information." *Proceedings of the 2003 Energy Services Conference.* Boca Raton, FL: Association of Energy Services Professionals.

Goldberg, Miriam, Oscar Bloch, Ralph Prahl, David Sumi, Bryan Ward, Rich Winch, and Tom Talerico. 2006. *Net-to-Gross Method Selection Framework for Evaluating Focus on Energy Programs*. Prepared for Wisconsin Focus on Energy.

Goldberg, Miriam, Rick Winch, Tom Talerico, Ralph Prahl, and Bryan Ward. 2008. *Treatment of Accelerated Savings.* Prepared for Public Service Commission of Wisconsin. KEMA, Glacier Consulting Group, Prahl & Associates, and PA Consulting Group.

Haeri, Hossein, Lauren Miller Gage, Matei Perussi, and Rick Ogle. 2005. *Evaluation of the 2002 and 2003 Transition Commercial Retrofit Programs*. Prepared for the Energy Trust of Oregon.

Hall, Nick. 2010. *Guidelines for State Conducting or Contracting Evaluations of ARRA Funded SEP Activities*. Prepared for US Department of Energy, Energy Efficiency and Renewable Energy Group, and Oak Ridge National Laboratory. TecMarket Works.

Hoefgen, Lynn, Lisa Wilson-Wright, Thomas Mauldin, and Tim Pettit. 2007. *NMR Memo to Sponsors of the MA Energy Star Appliances Program, Clothes Washer component: Estimates of Net Impact of the 2006 MA Energy Star Appliances Program, Clothes Washer component.* Cambridge, MA: NMR.

Hoefgen, Lynn. 2010. "Choosing the Right Tools: How Different Markets and Programs Call for Different Approaches to Estimating Net Savings." Paris, France: *Proceedings of the International Energy Program Evaluation Conference*, June.

Hoefgen, Lynn, Angela Li, Gail Azulay, Ralph Prahl, and Susan Oman. 2008. "Market Effects: Claim Them Now or Forever Hold Your Peace." Pacific Grove, CA: *Proceedings of the 2008 ACEEE Summer Study on Energy Efficiency in Buildings,* August.

Hoefgen, Lynn, Lisa Wilson-Wright, and Shel Feldman. 2004. "Assessing the Effects of Utility Programs on the Market Penetration of ENERGY STAR Appliances." *ACEEE-CEE Symposium on Market Transformation*, March.Hopper, Nicole, Charles Goldman, and Jeff Schlegel. 2006. *Energy Efficiency in Western Utility Resource Plans: Impacts on Regional Resource Assessment and Support for WGA Policies.* Ernest Orlando Lawrence Berkeley National Laboratory: Environmental Energy Technologies Division.

Holbrook, Allyson L. Jon A. Krosnick, and Allison Pfent. 2008. "The Causes and Consequenes of Response Rates in Surveys by the News Media and Government Contractor Survey Research Firms." In *Advances in Telephone Survey Methodology*, Edited by James M. Lepkowski, Clyde Tucker, J. Michael Brick, Edith de Leeuw, Lilli Japec, Paul J. Lavrakas, Michael W. Link, and Roberta L. Sangster. John Wiley & Sons, Inc.

Horowitz, Marvin J. 2008. "Energy Efficiency and Changes in Energy Demand Behavior." Demand Research LLC. Pacific Grove, CA: *Proceedings of the 2008 ACEEE Summer Study on Energy Efficiency in Buildings,* August.

Horowitz , Paul. 2009. *Glossary of Terms: Version 1*. A Project of the Regional Evaluation, Measurement, and Verification Forum. March.

Howell, Embry M. and Alshadye Yemane. 2006. "An Assessment of Evaluation Designs: Case Studies of 12 Large Federal Evaluations." *American Journal of Evaluation* 27(2):219-236.

Johnson, T. and M. Fendrich. 2005. "Modeling Sources of Self-report Bias in a Survey of Drug Use Epidemiology." *Annals of Epidemiology* 15(5): 381-389.

Jonas, Wayne. 2005. *Mosby's Dictionary of Complementary and Alternative Medicine*. Elsevier: Amsterdam.

Kahneman, Daniel K. and Robert Sugden. 2005. "Experienced Utility as a Standard of Policy Evaluation." *Environmental and Resource Economics* 32: 161-181.

Kandel, Adrienne. "Theory-Based Estimation of Energy Savings from DSM, Spillover, and Market Transformation Programs Using Survey and Billing Data." *Program Measurement and Evaluation* 10:149-162.

Keating, Kenneth M. 2009. "Freeridership Borscht: Don't salt the soup." Portland: *Energy Program Evaluation Conference*.

KEMA and Itron. 2009. *Final Draft: Scoping Study and Work Plan for a High Bay Lighting Market Effects Study.* Prepared for the California Public Utilities Commission.

Lenihan, Helena. 2004. "Evaluating Irish Industrial Policy in Terms of Deadweight and Displacement: A Quantitative Methodological Approach." *Applied Economics* 36.3: 229.

Loper, Joe, Steve Capanna, Rodney Sobin and Tom Simchak. 2010. *Energy Savings Credits: Are potential benefits being realized?* Alliance to Save Energy.

Mahone, D. and N. Hall 2010. "Pursuit of Aggressive Energy Savings Targets: Aligning Program Design and Evaluation." Pacific Grove, CA: *Proceedings of the 2010 ACEEE Summer Study on Energy Efficiency in Buildings,* August.

Maxwell, Jonathan, Cherie Gregoire, Jennifer Meissner, and Lori Megdal. 2009. "Large Lessons Learned: Impact Evaluation of Projects that Reported over 1,500,000 kWh/yr savings." Portland: *Energy Program Evaluation Conference*.

Mayne, John. 2001. "Addressing Attribution Through Contribution Analysis: Using Performance Measures Sensibly." *Canadian Journal of Program Evaluation* 16: 1-24.

Measurement, Analysis and Reporting Task Force. 2010. *Final Report of the Measurement, Analysis and Reporting Task Force.* British Columbia, Canada: British Columbia Utilities Commission.

Megdal, Lori, Yogesh Patil, Cherie Gregoire, Jennifer Meissner, and Kathryn Parlin. 2009. "Feasting at the Ultimate Enhanced Free-ridership Salad Bar." Portland: *Energy Program Evaluation Conference.*

Messenger, Mike, Ranjit Bharvirkar, Bill Golemboski, Charles A. Goldman, and Steven R. Schiller. 2010. *Review of Evaluation, Measurement and Verification Approaches Used to Estimate the Load Impacts and Effectiveness of Energy Efficiency Programs.* Ernest Orlando Lawrence Berkeley National Laboratory.

Mohr, Lawrence B. 1995. *Impact Analysis for Program Evaluation: Second Edition*. Thousand Oaks: Sage.

Mohr, Lawrence B. 1996. *The Causes of Human Behavior: Implications for Theory and Method in the Social Sciences*. Ann Arbor: University of Michigan Press.

Muncaster, Katherine, Erik Kaye, Sarah Smith, Murray Bond, Keith Veerman, and Jim Fraser. 2010. *Final Report of the Measurement, Analysis, and Reporting Task Force.*

National Action Plan for Energy Efficiency and EPA's State Climate and Energy Program. 2010. "Understanding Net Savings." Four presentations presented as part of the Webinar Series on Evaluation, Measurement, and Verification. Presenters: Steven Schiller, Jane Peters, Tim Drew, and Bill Saxonis.

Neugart, Michael. 2008. "Labor Market Policy Evaluation with ACE." *Journal of Economic Behavior and Organization* 67: 418-430.

Nexus Market Research. 2005. *Impact, Process, and Market Study of the Connecticut Appliance Retirement Program: Overall Report.*

Nexus Market Research. 2007. *Estimates of Net Impact of the 2006 Massachusetts ENERGY STAR Appliances Program, Clothes Washer Component.* Prepared for Cape Light Compact, National Grid, NSTAR Electric, Unitil, and Western Massachusetts Electric Company.

Nexus Market Research. 2009. *Phase I Report: Residential New Construction (Single-Family Homes) Market Effects Study.* Prepared for the California Public Utilities Commission. Published by KEMA.

Nexus Market Research, RLW Analytics, and Dorothy Conant. 2008. *Process Evaluation Report (MPER) for the 2007 Massachusetts ENERGY STAR® Lighting Program*. Prepared for Cape Light Compact, National Grid, NSTAR Electric, Unitil, and Western Massachusetts Electric Company.

Nexus Market Research, RLW Analytics, Shel Feldman Consulting, and Dorothy Conant. 2004. "Delphi Study." *Appendix H in Market Progress and Evaluation Report for the 2003 Massachusetts ENERGY STAR Appliances Program*. Prepared for the Massachusetts ENERGY STAR Appliances Program Administrators.

Nexus Market Research and Shel Feldman Consulting. 2005. "Statistical Analysis of Market Penetration of ENERGY STAR Compliant Appliances." Prepared for the Massachusetts ENERGY STAR Appliances Program Administrators.

NMR Group, Inc. 2010. *Results of the Multistate CFL Modeling Effort*. Prepared for Connecticut Energy Efficiency Board. http://www.ctsavesenergy.org/ecmb/documents.php?section=22.

Opinion Dynamics Corporation. 1997. *Wisconsin's Forced-Air Furnace Market: Tracking Residential & Small Commercial Sales.* Report Prepared for Energy Center of Wisconsin.

Opinion Dynamics Corporation. 2006. *Final Report: The New Hampshire Electric Utilities' Low-income Retrofit Program – Impact Evaluation.*

Oswald, Kim Johnson. 1991. Chicago: "Measurement Issues at the Free Luncheonette: Session Summary." *Energy Program Evaluation Conference*.

Peters, Jane and Marjorie McRae. 2008. "Free-Ridership Measurement is Out of Sync with Program Logic…or, We've Got the Structure Built, but What's its Foundation?" *Proceedings of the ACEEE Summer Study on Energy Efficiency in Buildings,* August.

Peters, Jane S. 2010. "Net to Gross Measurement: View from the Field." Presented at *US EPA Webinar—Understanding Net Savings,* April.

Peters, Jane, and Ryan Bliss. 2010. *Fast Feedback Pilot: Existing Buildings and Production Efficiency Programs*. Prepared for the Energy Trust of Oregon.

Peters J. S., M. Spahic, C. Jackson, and S. Lutzenhiser. 2010. "What are Consumers Thinking About Energy Today?" *Proceedings of the AESP National Energy Services Conference,* February.

Prahl, Ralph, Mariam Goldberg, David Sumi, Tom Talerico, Bobbi Tannenbaum, Bryan Ward, and Rick Winch. 2008. *Integrating Supply Side Results with End-User Net-to-Gross Self Reports.* Prepared for Wisconsin Focus on Energy. PA Consulting Group.

Rathbun, Pamela, Carol Sabo, and Bryan Zent. 2003. *Standardized Methods for Free-ridership and Spillover Evaluation—Task 5 Final Report.* Prepared for Focus on Energy and Public Service Commission of Wisconsin. PA Consulting Group.

Ridge, Richard, Phillipus Willems, Jennifer Fagan, and Katherine Randazzo. 2009. "The Origins of the Misunderstood and Occasionally Maligned Self-Report Approach to Estimating the Net-to-Gross Ratio." Portland:.*Energy Program Evaluation Conference*.

Rogers, Everett M. 2003. *Diffusion of Innovations* (5th ed.). New York: Free Press.

Rogers, Everett M. and F. Floyd Shoemaker. 1971. *Communication of Innovations*. New York: Free Press.

Rosenberg, Mitch and Lynn Hoefgen. 2009. *Market Effects and Market Transformation: Their Role in Energy Efficiency Program Design and Evaluation.* California Institute for Energy and Environment.

Rossi, Peter H., Mark W. Lipsey, and Howard E. Freeman. 2004. *Evaluation: A Systematic Approach.* Thousand Oaks, CA: Sage.

Rufo, Michael W. 1999. *Evaluation of the 1998 Nonresidential Standard Performance Contract Program*. Prepared for California Board of Energy Efficiency.

Rynes, Sara L., Barry Gerhart, and Kathleen A. Minette. 2004. "The Importance of Pay in Employee Motivation: Discrepancies between What People Say and What They Do." *Human Resource Management* 43: 381-394.

Saxonis, Bill. 1989. "Program Evaluation of the New York State Energy Efficient Appliance Rebate Demonstration Program: The Free Rider Factor." Chicago: *International Energy Program Evaluation Conference Proceedings*.

Saxonis, Bill. 1991a. "Free Riders and Other Factors that Affect Net Programs Impacts." In *Handbook of Evaluation of Utility DSM Programs*. Oak Ridge, TN: National Laboratory.

Saxonis, Bill. 1991b. "Measuring Free Riders: Does the Economic Climate Make a Difference? Measurement Issues at the Free Luncheonette." Chicago: *Energy Program Evaluation Conference*.

Saxonis, Bill. 1995. "After Ten Years of Measuring Free Riders: What Have We Learned? What Challenges Remain?" Chicago: *International Energy Program Evaluation Conference Proceedings.*

Saxonis, Bill. 2007a. "Free Ridership and Spillover: A Regulatory Dilemma." *International Energy Program Evaluation Conference Proceedings,* Chicago.

Saxonis, Bill. 2007b. "Energy Efficiency in New York: It Measures Up!" *ACEEE National*

Saxonis, Bill. 2010. "The Net Savings Dilemma." Presented at *US EPA Webinar—Introduction to Net and Gross Savings,* April.

Schauer, Laura. 2010. *Home Performance with ENERGY STAR: Insulation Supply-side Study Results and Integration with Participant Findings.* Prepared for Focus on Energy. PA Consulting Group.

Schlegel, Jeff, Miriam Goldberg, Jonathan Raab, Ralph Prahl, Marshall Keneipp, and Daniel Violette. 1997. *Evaluating Energy-Efficiency Programs in a Restructured Industry Environment: A Handbook for PUC Staff.* Prepared for National Association of Regulatory Utility Commissioners.

Scriven, Michael. 1976. "Maximizing the Power of Causal Investigations: The *Modus Operandi* Method." In Gen V. Glass, ed. *Evaluation Studies Review Annual Volume 1*. Beverly Hills: Sage.

Shadish, William R., Thomas D. Cook, and Laura C. Leviton (eds.). 1991. *Foundations of Program Evaluation: Theories of Practice.* Newbury Park, CA: Sage.

Skumatz, Lisa, M. Sami Khawaja, and Jane Colby. 2009. *Lessons Learned and Next Steps in Energy Efficiency Measurement and Attribution: Energy Savings, Net to Gross, Non-Energy Benefits, and Persistence of Energy Efficiency Behavior.* California Institute for Energy and Environment.

Spiegel, Alix. 2010. "The Growing Power of the Sugar Pill." *National Public Radio*. Story aired March 8; accessed June 25, 2010. Available at http://www.npr.org/templates/story/story.php?storyId=124367058&ft=3&f=124367058.

TecMarket Works. 2004. *The California Evaluation Framework*. Prepared for the California Public Utilities Commission and the Project Advisory Group, June.

TecMarket Works, et al. 2006. *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals.* San Francisco: California Public Utilities Commission.

TecMarket Works, et al. 2007. *Guidelines for Estimating Net-To-Gross Ratios Using the Self-Report Approaches.* California Public Utilities Commission Energy Division.

Titus, Elizabeth and Julie Michals. 2008. "Debating Net versus Gross Impacts in the Northeast: Policy and Programs Perspectives." Asilomar, CA: *Proceedings of the ACEEE 2008 Summer Study on Building Conference*.

Titus, Elizabeth, Monica Nexius, and Julie Michals. 2004. "How Do We Measure Market Effects? Counting the Ways, and Why it Matters." Asilomar, CA: *Proceedings of the ACEEE 2004 Summer Study on Energy Efficiency in Buildings.*

Trexler Climate Energy Services. 2006. *A Consumers' Guide to Retail Offset Providers*. Prepared for Clean Air Cool Planet.

Wilson-Wright, Lisa, Shel Feldman, and Lynn Hoefgen. 2004. "Can We Distinguish the Effects of Specific Program Support Activities on Market Penetration?" *Proceeding of the Annual Meeting of Association of Energy Service Professionals,* December.

Wilson-Wright, Lisa, Shel Feldman, Lynn Hoefgen, and Angela Li. 2005. "Front-loading Marketing: Assessing Cumulative Effects of ENERGY STAR Appliances Promotions on State-by-State Market Penetration Levels." *Proceedings of the Twelfth International Energy Program Evaluation Conference*, August.

Winch, Rich and Tom Talerico. 2007. *Comprehensive CFL Market Effects Study—Final Report.* Prepared for State of Wisconsin Department of Administration Division of Energy.

Winch, Rich, Tom Talerico, Bobbi Tannenbaum, and Pam Rathbun. 2008. *Framework for Self-Report Net-to-Gross (Attribution) Questions.* Prepared for Wisconsin Focus on Energy. PA Consulting Group.

Winch, Rich. and Tom Talerico of Glacier Consulting, Group, LLC. 2008. *Second Annual Comprehensive CFL Market Effects Study – Final Report*. Delivered to the State of Wisconsin Public Service Commission, September.

Windell, Phillip and Jane Peters. 1994. "Innovativeness, Predisposition and Freeridership." . Pacific Grove, CA: *Proceedings of the ACEEE 1994 Summer Study on Energy Efficiency in Buildings*, August.

Wobus, Nicole, Frank Stern, Brent Barkett, and Carole Nemore. 2009. "Attribution Analysis for Renewable Portfolio Standard (RPS) Programs: Challenges Encountered Evaluating New York's RPS." *Energy Program Evaluation Conference*, Portland.

World Resources Institute. 2005. *The Greenhouse Gas Protocol for Project Accounting.* http://www.ghgprotocol.org/standards/publications.

World Resources Institute. 2007. *Guidelines for Quantifying GHG Reductions from Grid-Connected Electricity Projects.* http://www.ghgprotocol.org/standards/publications.