BEFORE
THE PUBLIC UTILITIES COMMISSION OF OHIO

In the Matter of the Annual Verification of the Energy Efficiency and Peak Demand Reductions Achieved by the Electric Distribution Utilities Pursuant to R.C. 4928.66.	**)** **)** **)** **)** **)**	**Case No. 13-1027-EL-UNC**

REPLY COMMENTS OF AEP OHIO TO THE PUBLIC UTILITIES COMMISSION OF OHIO'S DECEMBER 11, 2013 ENTRY

INTRODUCTION

On December 11, 2013, the Public Utilities Commission of Ohio ("Commission" or "PUCO") issued an entry establishing a comment and reply period in order to assist the Commission in its review of Evergreen Economics' report ("2011 Evergreen Report" or "2011 Report") on the verification of the annual levels of energy efficiency and of peak demand reductions achieved by each of the electric utilities in 2011. The 2011 Report was filed on May 2, 2013, by Staff on behalf of Evergreen Economics ("State Wide Evaluator" or "SWE").

AEP Ohio ("Ohio Power Company" or "Company") appreciates the Commission's request for reply comments on the topics raised by parties in the initial comments submitted on January 13, 2014. Notwithstanding the Commission's order in Case No. 12-665-EL-UNC dated August 7, 2013, in which the Commission clearly stated that it is premature to address the issue of net savings in Ohio, the joint comments of the Environmental Law and Policy Center, Ohio Environmental Council, Sierra Club, and

Natural Resources Defense Council (collectively "Environmental Advocates"), continue to urge the Commission to transition to a net savings measurement standard in Ohio.

AEP Ohio offers these reply comments and states that its views or concerns could change depending on the facts and circumstances in the future. The purpose of these reply comments is an attempt to provide the Commission some specific and relevant input on the recommendations and issues raised in the initial comments. Further, the Company's failure to address any particular recommendation should not be construed as acceptance or agreement.

NET-TO-GROSS (NTG) ADJUSTMENTS

At page 2 of the Environmental Advocates' initial comments, they accept the premise in the 2011 Report that free ridership estimates based on the analysis of two questions for only two of the seven residential programs that AEP Ohio operates should be extrapolated to the whole without question. Further, the Environmental Advocates repeat the statement shown at page 6 of the 2011 Report, "[a]t some point in the future, however, the PUCO anticipates requiring the utilities to also report net impacts…"[1]. Yet, no reference for the statement is provided. The Commission order in Case No. 09-512-GE-UNC dated October 15, 2009, at pages 4-6, and the entry dated January 27, 2010, in the same case, at page 3-10 of the RFP, explicitly addressed the joint issues of free ridership and spillover/free drivers. Quite clearly, the Commission ordered that until expressly addressed by the Commission, the utilities will report gross savings. Further, these orders held that one aspect, such as free ridership, would not be addressed without

[1] Environmental Advocates initial comments at page 2.

the inclusion of all sides of the issues[2]. These decisions occurred after a lengthy public process to develop the rules governing energy efficiency and peak demand response in Case Nos. 08-888-EL-ORD and 09-512-GE-UNC.

The Environmental Advocates continue to assert that the time is now to adopt a net-to-gross framework. AEP Ohio respectfully disagrees. Contrary to the assertion that, "a transition to net savings in Ohio will be relatively simple"[3], it will not. Indeed, it will contribute to increased costs for evaluation at a time when AEP Ohio is trying to minimize overall program costs per kWh. Spillover analysis requires the use of a general population survey with sample sizes that are sufficient in number to find customers who installed or undertook measures that were available as part of the program, but who also did not participate in the program. These numbers are required regardless of whether or not the customer was a spillover participant.

AEP Ohio reasserts its initial comments on NTG estimates filed in Case No. 12-665-EL-UNC. Recent literature indicates that the issue of applying one-sided net-to-gross adjustments has come under considerable scrutiny in the three years since the Commission issued its order on using gross savings. Skumatz and Vine (2010, see Attachment 1) state:

> (T)he combination of the "negative" of free ridership and the "positive" of spillover are computed as a "net to gross" (NTG) ratio, and are applied to the "gross" savings to provide an estimate of attributable "net" savings for the program. The NTG ratio only equals free ridership (FR) if spillover (SO) is (or is assumed to be) zero.
>
> There is considerable – and growing - controversy regarding the use of net to gross, particularly in regulatory proceedings. As noted above, NTG ratios can be used to reduce (incorporating free ridership) or potentially expand (if spillover associated with the program exceeds free ridership)

[2] October 15, 2009 order in Case No. 09-512-GE-UNC at page 6, paragraph 16.
[3] Environmental Advocates initial comments at page 5.

> the amount of savings attributable to a program. The concern is that evaluations carefully estimate (gross) savings that were delivered, but then the savings (and, directly, the associated financial incentives to the agency delivering the program) are discounted by a free ridership factor measured by methods that are less "trusted" – in other words, specifically measuring gross savings based on statistical analysis of meter readings/ billing records, compared to measuring free ridership and/or spillover based on self-report surveys of hypothetical decisions and behavior.[4]

Wilson-Wright *et al* [5] (2011, see Attachment 2) discuss the challenges of evaluating free ridership in light of the multitude of influences and messages from the utility and other entities that can blur trying to determine whether the specific amount offered by a rebate program was the prime determinant of participation. AEP Ohio devotes a significant number of dollars to both program specific and general marketing campaigns in support of energy efficiency. Mahone's 2011 paper[6] uses a series of overlapping concentric circles to emphasize the impacts of an integrated portfolio of programs and marketing campaigns on the measurement of program impacts, free ridership, spillover, and market effects (see Attachment 3).

First, the law (R.C. 4928.66) established a compliance goal as a percent of retail sales measured against an average of the three preceding years as the baseline. It implicitly addresses the free ridership issue submitting mercantile projects performed during the baseline by determining that the retail sales of the baseline period will be

[4] Skumatz, LA, and Vine, Edward: A National Review of Best Practices and Issues in Attribution and Net-to-Gross: Results of the SERA/CIEE White Paper, Proceeding of the 2010 ACEEE Summer Study on Energy Efficiency in Buildings Volume5,p. 347

[5] Wilson-Wright,Lisa *et al*,Proceedings of the 2011 International Energy Program Evaluation Conference, The Intersection of Policy and Methodology in Net Savings Estimation: Recommendations from a Regional Scoping Study.

[6] Mahone, Douglas. 2011. Free-Ridership as a Way to Kill Programs – How Evaluation Policies Can Frustrate Efficiency Goals. Paper presented at the International Energy Program Evaluation Conference. Boston, MA.

adjusted upward to account for the reduced energy savings of the projects. The law is silent relating to making equivalent adjustments to the baseline for non-mercantile projects on a going forward basis.

Second, free ridership and how to estimate it accurately continues to remain an open research question. Estimates of free riders/free drivers/spillover are difficult to estimate with certainty given the reliance on survey questions asked months after the measures were installed.

Some evaluators have experimented with asking questions on rebate applications or time of sale documents. These too have their issues. One is limited in the number of questions that can be asked at the time of sale or on a rebate application due to time and/or space issues.

Third, compliance with goals requires a level of certainty that should not be muddied with after-the-fact adjustments. Applying estimates of past programs to current programs may be detrimental to compliance efforts. By the time free rider estimates are developed during the evaluation cycle, utilities are almost half way through their program year. The ability to make adjustments to program activity or program design is limited at that time.

Fourth, the literature indicates that after much effort and expenditures, it appears to be rare that large numbers of free riders are found when both free riders and spillover are considered.

CONCLUSION

AEP Ohio respectfully offers the preceding reply comments to assist the Commission in its review of the 2011 Evergreen Report as filed in this docket on May 2, 2013.

//s/ Steven T. Nourse
Steven T. Nourse
Matthew. J. Satterwhite
Yazen Alami
American Electric Power Service Corporation
1 Riverside Plaza
Columbus, Ohio 43215
Telephone: (614) 716-1608
Fax: (614) 716-2950
Email: stnourse@aep.com
mjsatterwhite@aep.com
yalami@aep.com

On Behalf of Ohio Power Company

CERTIFICATE OF SERVICE

I certify that a copy of AEP Ohio's foregoing **Reply Comments** was served by email upon parties of record on January 28, 2014.

//s/ Steven T. Nourse
Steven T. Nourse

A National Review of Best Practices and Issues in Attribution and Net-to-Gross: Results of the SERA/CIEE White Paper

Lisa A. Skumatz, Skumatz Economic Research Associates, Inc.
Edward Vine, California Institute for Energy and Environment

ABSTRACT

Energy efficiency evaluation / attribution methods have reached a point that they must evolve in order to provide credible evaluation results for the next generation of programs. Recognizing this need, a national review was undertaken to examine the state of the art, gaps, and next steps needed to meet the evaluation needs for new programs, including behavioral and educational initiatives.

This study used interviews, a literature review, and analysis from around the United States to examine technical, research, and policy issues associated with the attribution of savings to programs – including net-to-gross (NTG) ratios and its components, free ridership, spillover, and other issues. The project reviewed results of net-to-gross (and component) estimations from around the country to identify patterns in results for "categories" of programs, and examined best practices in net savings estimation methods used to date for traditional measure-based programs.

This study found considerable variation in NTG methods, coverage, and component results. This project also examined policies used by different states related to this topic, such as whether NTG or its components are used at all, whether "deemed" levels are used, or whether the regulators endorse or include NTG estimates based on primary research. Protocols from several states were reviewed and compared, and the strengths and weaknesses of the approaches were examined.

Beyond reviewing the "state of the art" in traditional attribution work, savings and NTG issues for behavior, education, and training-based programs were also analyzed. For these programs, savings are difficult to measure, and marketplace "chatter" and overlapping programs and deliverers make measurement especially challenging. Some areas of the country are specifically addressing issues related to errors in measurement associated with NTG, and these results are highlighted. Finally, the project examined gaps in existing research, promising techniques for non-measure-based programs, and recommended next steps.

Project Introduction / Context

On behalf of the California Public Utilities Commission (CPUC), this project sought to identify current and improved techniques – and associated policy issues – related to[1]:

- **Gross effects**: Measuring the broad array of impacts caused, or potentially caused, by program interventions – measure-based, market-based, education or other interventions. This includes the measurement of gross energy savings and non-energy impacts.

[1] This paper presents the findings from one of eight white papers on behavior and energy that were funded by the CPUC and managed by the California Institute for Energy and Environment (CIEE). This work does not necessarily represent the views of the CPUC or CIEE or any of its employees. The white papers are available at: http://uc-ciee.org/energyeff/energyeff.html.

- **Net effects attribution**: Identifying the share of those effects – direct and indirect – that can be attributed to the influence of the interventions undertaken – above and beyond what would have occurred without the intervention – either naturally or due to the sway of other market influences or trends.

The overall research examined four key topics in evaluation: gross savings; attribution / free ridership / net to gross (NTG); non-energy benefits; and persistence. This paper focuses on the second of these evaluation topics. The findings from these evaluation efforts play a critical role in an array of applications, from analysis to program design. Given that evaluation results are often used in making program and reward decisions that put significant investment dollars at risk, it becomes prudent to revisit methods and approaches. Further, as programs have evolved, evaluation has become more complex:

- Programs have moved away from "widget"-based programs toward behavioral, education, advertising, and upstream programs that make it harder to "count" impacts.
- There is an increasing number of actors delivering these programs – leading to market "chatter" and increasing difficulty in identifying which among all the deliverers of the energy efficiency "message" are responsible for the change in energy efficiency behaviors, actions, or purchases. The increased chatter in the marketplace creates a situation in which consumers may be influenced by any number of programs by local utilities as well as influences from outside the utility (national programs, neighboring programs, movies / media, etc.).

As a result, attributing or assigning responsibility for changed behaviors and the adoption of energy efficiency measures or services is muddied and challenging.

For this project,[2] SERA[3] reviewed more than 250 conference papers and reports, and reached out to 100 professional researchers for interviews to identify improved techniques (and associated policy issues) for quantifying the share of direct and indirect effects that can be attributed to the influence of program interventions above and beyond what would have occurred without the intervention – either naturally or due to the sway of other market influences or trends. The white paper addresses all four evaluation topics, but this conference paper focuses only on "net-to-gross" and its constituents, free ridership and spillover.[4]

The literature indicates that there are a number of uses to which free ridership, spillover, or NTG ratios are relevant. Free ridership helps to identify superior program designs and helps to identify program exit timing. Spillover helps to assess the performance of education / outreach

[2] The context for this paper (California) relates to, but is not exclusive to, the situation of programs run by utilities with oversight by a public service commission and where shareholder incentives are at stake and depend on the determination of attribution. This review has relevance beyond this situation, but readers in other states may need to make a few adjustments in terminology, etc.

[3] Skumatz Economic Research Associates (SERA) was commissioned by CIEE to conduct this review. The lead author wishes to thank the following for assistance in preparing the white paper: D. Juri Freeman, Dana D'Souza, and Dawn Bement (Skumatz Economic Research Associates), Carol Mulholland, Jamie Drakos, and Natalie Auer (Cadmus Group), and Gregg Eisenberg (Iron Mountain Consulting).

[4] This paper does not discuss "takeback". An example of takeback is when a homeowner turns up the thermostat after more efficient HVAC systems are installed. This review found little recent work on this topic.

/behavioral programs,[5] and it helps to identify program exit timing. Not examining free ridership and spillover *ex post* will make it impossible to distinguish and control for poorly designed / implemented programs, as well as for programs that may have declining performance over time and may have outlived their usefulness, at least in their current incarnation. Some interviewees said 'deemed savings are ridiculous' for this reason.

Definition and Methods – Net To Gross (NTG)

Identifying the "net" effects is a significant element of the assessment of benefits and costs for a program, computations that, in some states, can determine the start, continuation, or termination of a program's funding. Estimating the effects of the program above and beyond what would have happened without the program involves identifying the share of energy-efficient measures installed / purchased that would have been installed / purchased without the program's efforts. Some purchasers would have purchased the measure without the program's incentive or intervention. They are called "free riders" – they received the incentive but didn't need it. Others may hear about the benefits of the energy-efficient equipment and may install it even though they do not directly receive the program's incentives for those installations and are not recorded directly in the program's "count" of installations. This is called "spillover," and there are three types of spillover:

- Inside project spillover occurs, for example, when refrigerators are rebated, and the person receives / installs that equipment, and then later installs an energy-efficient dishwasher.
- Outside project spillover occurs, for example, when a builder receives rebates on one project, but installs similar efficient measures in other homes without rebates.
- Non-participant spillover occurs, for example, when a builder hears about energy efficiency and does not participate or receive any rebates, but decides to install efficient equipment to serve his customers or to keep up with other builders, etc. No incentives were provided for these measures.

Sometimes, the first two examples are referred to as Participant Spillover and the third example as Non-Participant Spillover.

The combination of the "negative" of free ridership and the "positive" of spillover are computed as a "net to gross" (NTG) ratio, and are applied to the "gross" savings to provide an estimate of attributable "net" savings for the program.[6] The NTG ratio only equals free ridership (FR) if spillover (SO) is (or is assumed to be) zero. The NTG, or its components, have been addressed in four main ways, described below. Each approach has pros and cons. We list key strengths and weaknesses of each method based on our literature review and interviews with evaluation professionals.

[5] For some of these types of programs, spillover is actually the point of the program, and omitting it ignores important program effects. Ignoring free ridership (in favor of "deemed" NTG figures) allows the continuation of poorly-designed or implemented programs, which wastes ratepayer money.

[6] The literature shows computations of this NTG ratio by adding the factors (1-FR+SO) or by multiplying the factors ((1-FR)*(1+SO)). Both are used in practice.

Deemed (Stipulated) NTG

A NTG ratio is assumed (1, 0.8, 0.7, etc.)[7] that is applied to all programs or all programs of specific types. This is generally negotiated between utilities and regulators or assigned by regulators.

- Advantages: Simple, uniform, and eliminates debate; no risk in program design or performance; inexpensive.
- Disadvantages: Does not recognize actual differences in performance from different programs, designs, or implementations.

NTG Adjusted by Models with Dynamic Baseline

A baseline of growth of adoption of efficient measures is developed, and the gross savings are adjusted by the changes in the baseline for the period.

- Advantages: Can reflect differences in performance for good or poor designs and implementation.
- Disadvantages: Complicated to identify appropriate baseline; data intensive; potentially expensive; introduces more risk to program designers related to program performance; may lead to protracted discussions.

Paired Comparisons NTG

Saturations (or changes in saturations) of equipment can be compared for the program (or "test") group versus a control group. The control group is similar to the test group but does not receive the program. Ideally, pre- and post- measurement is conducted in both test and control groups to allow strong "net" comparisons.

- Advantages: Can reflect differences in performance for good or poor designs and implementation; straightforward concept and reliable evaluation design.
- Disadvantages: Control groups can be difficult to obtain; if imperfect control groups are used, statistical corrections may be subject to protracted discussions.

Survey-Based NTG

A sophisticated battery of questions is asked about whether the participant would have purchased the measures or adopted the behavior without the influence of the program. Those participating despite the program are the free ridership percentage. These are then netted out of the gross savings. Spillover batteries can also be administered to samples of potential spillover groups (participants, non-participants).

- Advantages: Provides an estimate of free ridership and spillover; can explore causes and rationales.

[7] If the NTG is less than zero, then this reflects the likelihood of some free ridership.

- Disadvantages: Responses are self-reported leading to potential bias or recall issues; may be expensive; can be difficult to get good sample of respondents for free ridership; requires well-designed survey instrument which can be long and which affects response rate.

The measurement of spillover involves different issues than the measurement of free ridership. Free ridership emanates from the pool of identified program participants; the effects from spillover are not realized from the participating projects and, in many cases, not even the entities that participated. Identifying who to contact to explore the issue of spillover and associated indirect effects can be daunting.

Our interviews and literature review suggest that a number of states consider free ridership in the calculation of NTG, but do not include spillover in their analyses of program effects, such as California. This analytic asymmetry undervalues energy efficiency by incorporating only subtractions (such as free riders) from gross savings and ignoring potential additions (such as spillover).

Issues and Controversies in NTG Determination

There is considerable – and growing - controversy regarding the use of net to gross, particularly in regulatory proceedings. As noted above, NTG ratios can be used to reduce (incorporating free ridership) or potentially expand (if spillover associated with the program exceeds free ridership) the amount of savings attributable to a program. The concern is that evaluations carefully estimate (gross) savings that were delivered, but then the savings (and, directly, the associated financial incentives to the agency delivering the program) are discounted by a free ridership factor measured by methods that are less "trusted" – in other words, specifically measuring gross savings based on statistical analysis of meter readings/ billing records, compared to measuring free ridership and/or spillover based on self-report surveys of hypothetical decisions and behavior.

Another controversy relates to the fact that only a small minority of free ridership, spillover, or NTG studies report any confidence ranges, or even discussions of uncertainty. Until these issues are addressed, given the financial implications, it is unlikely much additional progress will be made in a more comprehensive treatment of free riders, spillover, or NTG in the regulatory realm. Furthermore, most behavioral and educational programs seem to be treated as indirect programs and not included in regulatory tests. This has a problematic side effect: lack of credits for benefits or savings from these programs results in an under-investment in these efforts. Because of their spillover implications, this puts educational (and potentially behavioral) programs at a disadvantage in portfolio development, designing rewards and incentives, and in resource supply applications.

In some states (e.g., California), these measurements have huge potential financial impacts in which utilities may receive financial awards for running programs and running them well. Based on the interviews and research, the controversy seems to arise from the following main sources:

- The potential for error and uncertainty associated with these measurements, because of difficulties in (1) identifying an accurate baseline; (2) identifying and implementing a control group; or (3) relying on self responses to a survey.

- The expense of high quality analysis – with arguments that the money could be better spent on program design, implementation, incentives, etc.
- Baselines and effects are harder and harder to identify and analyze as programs move up stream, involve different levels of vendors and other actors, and lead to changes in baselines up the chain. In addition, program spillover complicates the identification of a reasonable control or comparison group.
- The difficulty in separating out the effects and influences of different programs within a marketplace (own utility / agency and outside utility / agency), often called "chatter".
- Concerns that using measured NTG or free ridership ratios introduces a great deal (to some, an unacceptable level) of risk or uncertainty into the potential financial performance metrics for the program, which will lead to "same old / same old" programs and reduce innovation in program offerings. [8]

Baselines are a very important part of the problem of measuring NTG, free ridership, and spillover. The calculation of baselines is complicated by several factors, including the difference between prescribed and actual practice, and the challenge of documenting what has not happened. Baselines relate to what would have happened without the program, which is generally understood to mean standard practice. Standard practice might generally be expected to relate to codes and standards, but this is not necessarily the case. In one study (referred to in Mahone 2008), the issue of baseline was found to be quite complex. Mahone (2008) notes that for at least the multifamily sector, none of the buildings were being built to the level of baseline codes – i.e., they were underperforming, so that the actual baseline of standard practice was below the baseline of codes. In this case, NTG would be estimated as greater than "one," since the energy efficiency program improved performance over the standard practice baseline.

Documenting what "would have happened" is the biggest challenge in evaluation (Saxonis 2007). Many interviewees suggested that strong market assessment is needed up-front to provide the maximum amount of baseline information. However, when it comes to the dynamic retail sector, it may be impossible to predict what they would have done without the program (Messenger 2009) – especially if changes occur upstream.[9] More research on standard practice in the field would provide a stronger basis for baselines and provide a sounder basis for determining NTG ratios.

What Precision Is Needed?

Assuming part of the concern about NTG relates to the accuracy of its computations, two questions arise before either including or excluding NTG – and specifically free ridership - across the board. First, how accurate does the NTG need to be for different possible applications, and second, are there computation approaches that provide that – or those varying – degree(s) of accuracy?

[8] Innovation is valuable, but agencies will not innovate (cannot justify innovating) in programs unless the risk is reasonably predictable. However, on the other side, regulators must assure that the reward structure doesn't encourage ineffective programs and that funding is spent appropriately and prudently.

[9] For example, some upstream changes may spill over to areas that might otherwise be considered potential control areas. If a manufacturer is induced to change the manufacture or mix of product, and they do so for California which is a big enough market to swing production in general, then the new product lines will become available in the potential control areas and the (important) market effect is then reduced.

The 2003 Nobel-award winning economist, W.J. Granger, noted that evaluations should be designed to the level of 'helping *avoid making wrong decisions (about programs)*'. The evaluation industry also makes a pertinent point that things that are measured tend to improve. Evaluators want to make sure that the following right decisions are made:

1) Assure public dollars are being responsibly spent;
2) Apportion dollars and efforts between alternative strategies; and
3) Help to identify the appropriate time for exit strategies (or program revisions).

This overriding principle has implications relevant to standards for evaluation in energy efficiency. It implies that the level of accuracy applied to evaluation research can be flexible, based on the value (cost) of the possibility of a wrong decision coming out of the particular advisory research. For example, making a decision on going ahead with a program or intervention may allow a much less accurate estimate for input information than a decision about the precise level of shareholder dollars that should be allowed for a particular agency. Thus, it is important to see how NTG results will be used, such as in the following activities:

- **Program planning**: Providing estimates of savings attributable to a program that can be used for program planning purposes (e.g., cost-benefit data).
- **Program marketing and optimization**: Providing quantitative feedback that helps to inform the design, delivery, marketing, or targeting of programs, including revisions to incentives, outreach, exit timing, or other feedback. The evaluation information can be used to understand tradeoffs, benefit-cost analysis, and decision making.
- **Integrated planning, portfolio optimization, and scenario analysis**: Providing savings and other feedback across and between programs that helps optimize program portfolios.
- **Generation alternative**: Providing an estimate of energy savings attributable to a program which may support a decision in deferring new generation.[10]
- **Performance incentives**: Providing estimates of savings attributable to a program that may be used to compute incentives to various agencies in return for efforts in program design, implementation, and delivery.

The degree of accuracy needed in the NTG computation for these various applications are more stringent (higher) if higher dollars are involved, e.g., if shareholder incentives are involved, or if a new power supply is being sought. The accuracy needed to avoid making a wrong decision varies directly with the potential dollars associated with that wrong decision. To illustrate the point, consider the following. "One size fits all" policies are perhaps not the best approach for including or excluding spillover in NTG computations. Ignoring spillover (because we are concerned that the accuracy of the estimates is of concern) for a program for which spillover is a key goal and outcome increases the chances of making a "wrong decision" about that program investment – and eliminates the chance to improve that performance (assuming measurement breeds improvement). Estimating spillover and applying ranges or confidence intervals to the

[10] For example, if a high amount of savings or value is assigned to the program.

values in assessing the program[11] may be preferable to ignoring spillover. On the other hand, ignoring spillover for a low value program or for a program for which spillover is not an integral part may not be a significant concern.

NTG Practices, Results, and Patterns

Several states use the California Standard Practice Manual, or large portions of it, for estimating energy savings, free ridership, non-energy benefits, and benefit-cost regulatory tests, including Oregon, Washington, Idaho, Montana, Wyoming, Utah[12], Iowa, Kansas, Missouri, New Mexico, and Colorado (Hedman, 2009). Several studies specifically examined state and utility practices regarding free ridership and net-to-gross. These studies find that utilities treat the issue of NTG differently. In some cases, there is no regulatory agreement on the estimation of NTG, and they historically treat free ridership only in the calculation of the NTG ratio. The Nevada Power and Sierra Pacific Power collaborative examined free ridership and spillover in 23 states and/or utilities serving states. They found 15 states (69%) did not use free ridership in estimating net savings (Quantec 2008). Other states say NTG is too costly and biased. Massachusetts prefers to have utilities focus on market transformation programs and correct for factors affecting NTG savings in program design. California requires deemed free ridership values in the calculation of the NTG, but excludes spillover. Several other states say estimating NTG is not a priority - they feel free ridership is balanced by spillover and make no further efforts, argue that measurement of free ridership and spillover is unreliable, or say that when they did measure it the value was close to one.

In Illinois, NTG ratios of 0.8 are assumed for low income programs and are lower for appliance efficiency programs (Baker 2008). Washington reportedly doesn't support savings from behavioral changes or NTG allowances or disallowances (Drakos 2009).

In addition to studies reviewing state and regulatory practices or guidelines, this project also examined patterns in NTG values, results, or methods across programs and regions. The authors assembled and reviewed more than 80 evaluation studies from California, New England, and the Midwest that contained estimates of free ridership and/or other elements of NTG. The studies, which covered residential (including low income) and commercial programs, provided estimates for lighting, HVAC, new construction, appliances, motors, and other measures delivered through incentive and non-incentive programs. The studies covered programs dating from 1991 to 2008. The project examined the studies for patterns in methods between areas of the country, and in free ridership and NTG results by sector, measure, or region. Although the studies were assembled as a convenience sample, and not a statistical sample, we found the following general results, methods, and gaps presented in Table 2.

Measure-level NTG performance varied, presumably depending on elements of the underlying program design and possibly due to measurement techniques as well. While these findings are useful, additional, and more comprehensive, work of this type is clearly needed before broad conclusions can be drawn.

[11] Or looking for that threshold value of spillover that "turns the decision" may be another way to address the accuracy issue. If the threshold is outside the estimated range for spillover or outside any credible or feasible range based on the rough estimate, the program decisionmaking is improved.
[12] Utah only allows one year of lost revenues in the Rate Impact Test.

Table 2: NTG Results

	Net To Gross , Free Ridership, Spillover
General results	• Most utilities and regulators exclude NTG or assume values that incorporate only free riders and range from about 0.7 to 1.0 (*ex ante*). *Ex post* results have been measured for many programs; spillover is measured much less often than free ridership (and spillover is more commonly reported in the Northeast than in California). • Most studies rely on self-report surveys using variations in questions incorporating partial free ridership/likelihoods; only a small percent used logit/ranking/discrete choice modeling. • Some studies included both *ex ante* and *ex post* NTG figures for the same program. The *ex post* values were generally 10-20% lower than the *ex ante* values. The most obvious exceptions were some cooking measure programs (*ex post* was about half the *ex ante* value), and some refrigerator programs that reported spillover values greater than 0.5. • Gaps included: Fewer than 10% reported confidence intervals; only a small subset covered NTG for gas savings; and very few studies identified free ridership for electricity savings; most considered only kWh effects.
Variations by measure type, program type or region	• Clear patterns for free ridership, spillover, or NTG results by measures, program types, and regions have not been demonstrated to date. The assumption is that variations in specific program design and measure eligibility definitions are important to results. NTG results in the literature are also affected by whether or not spillover is included in the assessment. • *Ex-post* free ridership clustered around 0.1-0.3 but ranged as high as 0.5 to 0.7 for some commercial HVAC / motors and refrigerator initiatives. *Ex-post* NTG clustered around 0.7-1.0, but dipped as low as 0.3 and as high as 1.3. The lowest free ridership was low income programs (as low as 0.03). • NTG for whole homes and home retrofits tended to be high (0.85 to 0.95), but ranged from 0.5 to more than 1.0. • Net realization rates were provided for about one-third of the programs, and the values averaged about 0.7 to 1.0. A number of values exceeded 1.0, including commercial HVAC rebate programs (1.07) and refrigerator rebate programs (1.15). Several programs showed net realization rates between 0.3 and 0.5 including several CFL programs, some refrigerator programs, some gas cooktop rebate programs, and some energy management system initiatives.
Variations for behavioral vs. measure-based programs	• Studies addressing NTG, free ridership, or spillover estimates associated with strictly behavioral programs were not found, and if available, are probably too few in number to lead to overarching conclusions or patterns.

Emerging Methods and Recommendations

Based on this project's analysis of the literature and interviews with evaluation professionals, the following findings and recommendations regarding NTG determination are presented:

- **Incorporate the refinements made in standard practices**. Historically, fairly simplistic measurement methods have been used to estimate free ridership. The computations have been based on self-reports. Sources of error with this method stem from faulty recall, bias toward claiming the program was not influential or influential, and from bias introduced in the form of hypothetical questions.

  The literature review noted improvements in self-report methodology including questions to distinguish "partial" free ridership. Later, studies combined partial free ridership with a review of "influencing factors" or "corroborating questions" which were used to adjust free ridership reports based on the combined evidence from the other

questions. For example, the questions might ask about the importance of the rebate in decision-making, whether the purchase was moved forward two years or more, whether they were already aware of the measures, and similar questions, and used these responses to validate or adjust responses to direct free ridership responses (Skumatz, Woods, and Violette 2004).

Other approaches have established multiple criteria for free ridership. In one study, free riders had to meet four criteria: aware of the measure before the program, intending to purchase before the program, aware of where to purchase the measure, and willing to pay full price. If the four conditions were met, the household or business was classified as a free rider. In another example, the Energy Trust of Oregon conducts long-term tracking on a number of programs –they assess the market, identify program influencers, and conduct in-depth research in order to determine how much of the gross savings to claim for the programs (Gordon 2008).

- **Recognize we may need to allow "credit-splitting or credit-sharing".** One key refinement may be the recognition that we may not be able to attribute "causality" to one program or intervention, but may need to consider splitting the credit. The issue of "chatter in the marketplace" is a concern, but this is also an issue for technology / measure / economic based programs as well as education / outreach programs. However, the industry has been more willing to apply causality to technology measures because we can see something put an implementation or desired decision "over the top" more clearly. It is important to understand what is happening in the market and if a 0/1 litmus test is required for causality, it is unlikely to be "proved" as attributable to a particular program or element (Messenger 2008). Recent attitudinal research from the Energy Center of Wisconsin confirmed that people get energy-saving information from multiple sources and concluded that… "it may take a village to raise a behavioral kilowatt-hour sometimes" (Bensch 2009). This may make it hard to attribute the kilowatt-hour to one specific influencer, but that doesn't make the kilowatt-hour less real or mean that the program had zero effect. The solution may be to acknowledge shares of the kilowatt-hour to multiple contributing factors (for behavioral and technology measures) and share the credit (Bensch 2009). And sharing the credit may be the right answer, as people may only pay attention if it is a 'whole choir singing the "save energy" song' (Bensch 2009). Sulyma (2009) argues that it is more than time to move beyond only "one" plausible explanation for impacts, and that probabilistic methods should be used to address this attribution issue.

- **Require random assignment for participants and non-participants for as many program types as feasible**. The experimental design approach has been well known for decades, with random assignment of eligible participants assigned to treatment and non-treatment groups. This helps address the baseline issue in a credible way. However, to implement this option would require the regulators, utilities, or agencies to "bite the bullet" in terms of the political fallout from those that want to participate but are put into the "no treatment" bucket. Or future participants could be put "on hold" – they could be used as a control group in the short term, but can participate in the program at a later time. This approach may be especially important for outreach and behavioral programs. Train (2009) suggests pairing this with a discrete choice model to predict behavior.

- Many interviewees also agreed that well-designed randomized control and treatment groups are well-suited to impact evaluation (and attribution) for behavioral programs; however, the evaluators and regulators have not developed the kind of faith in them that they have in other programs. The use of these approaches with appropriate modeling (including mixed logit, discrete choice, etc.) shows promise (Ridge et. al. 2009, Train 2009). There is also concern that these random techniques may become more complicated, as controlling for the many influences is complex (including spillover), making a battery of questions important to the analysis (Messenger 2008, Cooney 2008, Train 2009). However, these kinds of tools – well-accepted in other social fields and with history in energy - apply well to energy-based behavioral programs. More evaluations of behavioral programs, and greater widespread cataloguing of the results (along with time), may be necessary to gain greater acceptance by regulators.

- **Consider survey designs that introduce a real-time data collection element.** There have been several instances in which utilities have introduced NTG-surveys as part of the program participation documents and gather early feedback – near the point of actual decision-making – on the program's influence in adopting the measures (Gordon and Skumatz 2007). This provides several benefits: increases return rate / sample size (and eliminates the problem of finding participants after they have moved or after years of delay); provides on-going data and allows evaluation at virtually any point after the program is implemented to support on-going refinement of programs; significantly reduces the cost of surveying and evaluation; provides more accurate data if the point of feedback is close to decision-making (recall may be improved); and helps to sort out which programs had what degree of influence. This may be suited to education and behavioral programs as well as "widget" programs, but needs testing, as the approach has not been widely applied.[13]

- **Consider discrete choice modeling approaches.** These approaches introduce explanatory variables that help to address issues of imperfect control groups, unobserved factors, etc. to allow improved estimates of attributable impacts. A discrete choice model predicts a decision made by an individual (purchase a measure, adopt a behavior, participate in a program) as a function of a number of variables, including demographic, attitudinal, economic, programmatic, and other factors. The model can be used to estimate the total number of eligible households, businesses, etc. that change their behavior in response to a program or action. The model can also be used to derive elasticities, i.e., the percent change in participation or behavior change in response to a given change in any particular (program design, demographic, or other) variable.

- **Consider compromise or "hybrid" approaches for fiscal-related applications.** A case might be made that the most "accurate" metric is pure *ex-post* measurement especially when those estimates are used for planning and reward purposes. If the main "rub" arises when NTG elements are part of the computations of financial reward or program approval, there are several possible options for the short term (until a "grander" solution is identified). Short-term deemed values (1-2 years of a new program that differs from

[13] It has been suggested that the smart grid or technologies might enhance the opportunity for real time collection of some important data elements.

traditional offerings) could be identified, allowing time for development and refinement of new, creative programs without punishing fiscal consequences. The program could be dropped if performance doesn't meet the offerer's expectations, and the method avoids an innovation penalty. True-up at some point is necessary to assure that the field learns about the performance of different types of programs and to assure that ineffective programs are not rewarded indefinitely. Deemed spillover values may be especially needed for programs targeted at education. Long-term deemed values could be allowed for well-known program types based on measured NTG from programs around the nation, where program performance is checked every 3 years, and where programs are penalized that perform more poorly than the norm, or require program comparisons against "best practices" periodically (every 3 or so years). Again, periodic true-up is needed. Another "tweak" to test to encourage innovation might be allowing differential rewards: upside incentives could potentially be larger than downside penalties for innovative programs. For some large, important, or innovative programs, negotiations for a priori values might be used.[14] Fiscal incentives must encourage (or at least not penalize) innovation, or only mediocre or "same old" programs will be offered – and they will be offered well past when they should be out of the market.

Reliable measurement methods are available that suit many program types, but more work remains, including research needs in the following areas: [15]

- Greater application of enhanced NTG, free ridership, and spillover methods incorporating partial (and/or deferred) free ridership and corroborating information.
- Greater use of experimental design (including random assignment for participants and non-participants) for as many program types as feasible.
- Comprehensive market assessment work for baseline support, on non-participant spillover, and modeling of decision-making. This is particularly important for many training, education, and behavioral programs.
- Data collection approaches that introduce a real-time data collection element piggybacking on program handouts / materials / forms and to allow periodic reviews of performance in time to refine programs.
- Discrete choice and other modeling methods, and statistical techniques to help address issues of imperfect control groups, unobserved factors, etc., to allow for improved estimates of attributable impacts.
- Accumulation of results on elements of NTG in a database and continuously updated with new research and evaluations, so comparisons and tracking are facilitated.

[14] This may cover programs such as those offered to only a very few large businesses (industrial, etc.), for example. This is suggested by the method NYSERDA is implementing for measuring NTG from their custom program that has very few participants (Cook 2008).

[15] And, as recognized by one of the paper's reviewers, these "methods-type recommendations" do not touch on issues such as who does the evaluation and the ability to share results for real-time program improvement.

Summary

Estimating the effects of the program above and beyond what would have happened without the program involves a relatively complicated step – identifying the share of energy-efficient measures installed / purchased that would have been installed / purchased without the program's efforts. Traditional elements include free ridership and spillover, combined into a NTG ratio. Spillover is more complicated than free ridership to measure, and as a consequence, a number of utilities that include free ridership never estimate spillover. However, given that many of the benefits from outreach and educational programs – and from a host of "non-widget-based programs – are realized from "spreading the word" (and the behaviors that follow), developing and using reliable and trusted methods that incorporate free ridership in program computations is a priority. These results are needed for applications including program design / assessment / refinement / portfolio development, program exit timing, and incentives.

Reasonable reliability is needed to provide useful information. To provide the best chance for optimal programs, several things are needed. NTG, free ridership and spillover estimates that are as reliable and precise as needed for the particular use – with greater precision needed for the calculation of program or portfolio incentives vs. quasi-quantitative / qualitative uses. NTG, free ridership and spillover estimates that provide replicable results and are based on credible, defensible estimation methods suited to the accuracy needed are a critical step in getting NTG results included in design and evaluation. Methods suited to different levels of accuracy for estimates of NTG, free ridership and spillover at reasonable cost levels would help optimize expenditures where they are most needed, and balance the tradeoffs of program funds vs. evaluation expenditures. Similarly, there should be flexibility in the application of NTG, free ridership and spillover results depending on type of program (whether programs are new / innovative / pilot; "same-old-same-old"; cookie cutter; custom; information-based; etc.).

Finally, it is critical that the application of NTG results is conducted in ways that avoid discouraging the development of new and creative and potentially effective programs. NTG should be applied in ways that properly assess program performance, but makes the risk of fiscal investment in (especially, new and innovative) programs manageable and reasonably predictable.

Current incentive structures, calculating attribution among actors, and the difficulty in identifying "participants" in new programs are discouraging innovation and leading researchers to consider discarding NTG analyses as a tool in energy efficiency evaluation. This is throwing the baby out with the bathwater. Instead, more widespread application of some of the approaches summarized in this paper can preserve the positives but not be hampered by the negatives of traditional NTG assessment. These evaluations are needed to "help avoid making a wrong decision…" with the public's money. To do this effectively, we need good methods, and we need to make sure the results are fed back into programs to be used in decision-making.

References[16]

Baker, David S., 2008. Illinois Department of Commerce and Economic Opportunity, Energy Division. Interview with the author, November 20, 2008.

Bensch, Ingo, 3/31/09. Energy Center of Wisconsin, WI, Personal Communication with the Author.

Cooney, Kevin. 7/31/08. Summit Blue, Boulder, CO, Personal Communication with the Author.

Drakos, Jamie, 3/31/09. Cadmus Group, Portland, OR, Personal Communication with the Author.

Gordon, Fred, 7/25/08. Oregon Trust, Portland, OR, Personal Communication with the Author.

Hedman, Brian. 3/31/09. Cadmus Group Portland, OR, Personal Communication with the Author.

Mahone, Douglas. 2008. Email from Douglas Mahone to Robert Kasman (PG&E), June 12, 2008, provided to author.

Messenger, Michael. 7/29/08 and 4/3/09. Itron, Sacramento, CA, Personal Communication with the Author.

Quantec, Scott Dimetrosky. 2008. "**Assessment of Energy and Capacity Savings Potential in Iow**a", Prepared for the Iowa Utility Association, February 15, 2008.

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**", *Proceedings of the IEPEC Conference*.

Saxonis, William. 2007. "**Free Ridership and Spillover: A Regulatory Dilemma**", *Proceedings of the IEPEC Conference*.

Gordon, Susie, and Lisa A. Skumatz, 2007, "**Integrated, Real Time (IRT), On-Going Data Collection For Evaluation – Benefits And Comparative Results** (Gordon & Skumatz, 4323), Proceed*ings for the European Council for an Energy Efficient Economy (ECEEE)*, France, June.

Skumatz, Lisa A., 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**", prepared for CIEE, December. Available at: http://uc-ciee.org/energyeff/energyeff.html.

Skumatz, Lisa, Dan Violette, and Rose A. Woods, 2004. "**Successful Techniques for Identifying, Measuring, and Attributing Casualty in Efficiency and Transformation Programs**." *Proceedings of ACEEE Conference*, Asilomar, CA.

[16] The long list of references underlying this project is cited in Skumatz 2009, the full report from which this article is excerpted.

Sulyma, Iris M., Power Smart, BC Hydro, Vancouver, Canada, Personal Communication with the Author, 4/28/09.

Train, Kenneth, UC Berkeley and N/E/R/A, CA, Personal Communication with the Author. 10/6/09.

The Intersection of Policy and Methodology in Net Savings Estimation: Recommendations from a Regional Scoping Study

Lisa Wilson-Wright, NMR, Melrose, MA
Lynn Hoefgen, NMR, Somerville, MA
Jane Peters, Research Into Action, Portland, OR
Marjorie McCrae, Research Into Action, Portland, OR
Elizabeth Titus, NEEP, Lexington, MA
Cheryl Browne, NMR, Somerville, MA

ABSTRACT

Energy efficiency community members are far from unanimous in their understanding and use of the concept of net savings. To work toward consistency and develop a research agenda, the Regional Evaluation, Measurement, and Verification Forum (Forum) facilitated by Northeast Energy Efficiency Partnerships (NEEP) commissioned a scoping study consisting of a literature review, interviews with 12 energy efficiency and air regulation experts, and feedback from Forum members.

The study confirmed that the challenges surrounding the definition and measurement of net savings persist more than 20 years after they were first enumerated. Such challenges include defining net savings, accessing reliable data, and measuring the counterfactual, or what would have happened in the absence of the program.

The study also identified new challenges, such as the issue of attribution, in the sense of separating the effects of a given program from the effects of all other programs or pro-efficiency messages to which participants have been exposed, at a time of rapid increase in the number, scope, and diversity of energy efficiency programs. Another new challenge is the added expectation by energy and air regulators in some parts of the country that energy efficiency programs help meet ambitious targets for reducing greenhouse gas emissions; the study shows that the air regulation and energy efficiency communities do not share a common understanding of how to translate energy savings into emissions reductions.

The paper provides a summary of these issues and presents policy and research recommendations to address the challenges laid out in the scoping study.

Introduction

The energy efficiency community pursues the estimation of net savings in order to separate the energy savings directly caused by a ratepayer or publicly funded program from those savings that would have happened anyway, and thereby determine whether the program has used the funds wisely. Estimating net savings, however, necessitates estimating a counterfactual. This charge is extremely difficult—some would say impossible—to fulfill, and the results always embody a degree of uncertainty because one can never know for certain what would have happened without the program. Despite the challenges of measurement, estimating net savings is critical to energy efficiency program assessment; without such estimation one cannot be sure that a program or portfolio has an effect on energy savings. 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 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. It is drawn from a larger scoping study completed by the NMR Group, Inc. (NMR) and Research Into Action, Inc. (RIA) (2010) for the Regional Evaluation, Measurement, and Verification Forum (Forum), which is a stakeholder group that includes energy and environmental regulators and energy efficiency providers in New England, New York and the mid-Atlantic, and is facilitated by Northeast Energy Efficiency Partnerships (NEEP).[1] The Forum designated a project subcommittee consisting of Forum members to assist NEEP and the evaluators. The subcommittee and evaluators worked together to identify potential interviewees and provide input into the project scope, direction, and conclusions. Although the Forum commissioned the study, members of the energy efficiency community throughout North America wrestle daily with these issues, and it is the authors' belief that the findings and recommendations apply beyond the Northeast.

Background

The Forum had three motivations for pursuing this study. The first motivation 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 estimate net savings, but each methodology suffers from some reliability, validity, and bias concerns. The lack of consistency in definitions and methodologies presents challenges for inter-jurisdictional efforts to meet energy and greenhouse gas emission reduction goals, for program administrators operating in multiple jurisdictions, and for market-level programs expected to have effects beyond the jurisdiction in which they operate. The Forum sought to assess the possibility of developing a consistent regional approach to net savings, and to develop an understanding of what this approach might look like if it were to be developed and adopted across the Northeast.

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 recently been expanded, and energy efficiency programs are expected to provide substantial reductions in greenhouse gas emissions. The Forum was interested in understanding the extent to which current ways of defining and measuring net savings meet the needs of these diversifying audiences.

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. Increasingly, however, when the energy efficiency community mentions "sorting out attribution," it refers to the fact that reductions in end users' energy consumption can be affected by myriad efficiency programs offered by a broad range of sponsors as well as by the economic, social, and technological context in which programs operate. This situation has significantly exacerbated the difficulty of establishing causation and estimating net savings. The Forum wanted to explore how the increasing challenges associated with attribution may affect net savings definitions and methodologies.

Scope

The scoping study drew distinctions between "adjusted gross savings," which include directly observable adjustments to gross impacts such as measure persistence and in-service rates, and "net savings," which include adjustments to gross impacts due to program attribution or customer behavior

[1] The Forum was established in 2009. See www.neep.org/emv-forum for a list of Forum members and a full description of Forum goals and products.

such as free-ridership and spillover. This paper limits its scope to discussions of net energy savings only. The authors acknowledge that program impacts go far beyond energy savings to include a wide range of non-energy impacts. In fact, the potential impacts are so numerous that the Forum directed the authors to focus only on net savings and non-energy impacts that may be calculated directly from net savings, such as reductions in greenhouse gas emissions,. It is important to note that many of the experts interviewed for this paper singled out the *regulatory* emphasis on net savings over other program impacts as a factor that exacerbates the challenges associated with net savings.

Methodology

The authors 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 provides a context—over time, across locations, and beyond energy efficiency—for issues related to program impacts generally and net savings in particular. In conducting the review, the authors 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.

The authors 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.[2] While the literature review provided the authors with an understanding of the dominant concerns, issues, and viewpoints related to net savings, the in-depth interviews allowed them to focus more specifically on the issues of greatest concern to the Forum that had not been adequately addressed in the existing literature. The results also incorporate comments made by Forum members and other individuals whom NEEP asked to review the paper. The authors treat such comments in the same manner as information gathered from the in-depth interviews. Throughout the discussion the authors use "commentators" to mean authors of the literature, interviewees, or reviewers of the scoping paper.

The authors analyzed the information gathered from the literature review and interviews using standard qualitative analysis techniques. Specifically, they organized the material according to the various themes of most concern to the Forum, such as definitions of net savings, estimation methodologies, intended use of the results, and strengths and weaknesses of approaches, among others. Based on this organization, they identified patterns and connections among ideas, which served as the basis for the qualitative discussion presented in the scoping paper and summarized here.

Results

This paper focuses on the recommendations and conclusions from the scoping paper completed for the Forum (NMR & RIA 2010). However, the authors find it important to summarize the key findings in order to provide the background necessary to understand the recommendations and conclusions.[3]

---

[2] 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, the authors did not include system planners among the interviewees. The authors 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.

[3] Note that the full scoping paper specifically addresses arguments for and against measuring net savings and the advantages and disadvantages of current methods of measuring net savings. The authors have decided not to address these topics in the

Context Surrounding Net Savings Estimation

The authors identified three contextual issues that greatly influenced the results, conclusions, and recommendations of the scoping paper: 1) key audiences for net savings, 2) different conceptualizations of net savings, 3) estimation of what would have happened in the absence of the program, which is a counterfactual.

Key Audiences for Net Savings. Historically, program administrators, energy regulators, and program planners have been the main audiences for net savings estimates. 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 (Friedmann 2007; Messenger *et al*. 2010; Saxonis 2007, 2010). 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 in response to increased funding and expanded regional goals for energy savings and reduced emissions. **Table 1** on the next page lists the audiences for net savings estimates identified through the research, specifies whether the audience is well established or emerging, and lists the audience's chief uses of net savings estimates. As the table shows, the established audiences for net savings tend to use the estimates for similar purposes: assessing if the program has achieved its goals and determining if the program has used funds wisely. In contrast, the primary emerging audience—the air regulation community—has conflicting views about how they will determine the impact of efficiency programs on the reduction of greenhouse gas emissions. It remains unclear whether the air regulation community will prefer to use energy savings in the form of gross savings, adjusted gross savings, or net savings, but their decision could have substantial effects on the estimation of net savings in the near future.

Conceptualizations of Net Savings. The second contextual issue involves how *net savings* is conceptualized. Sources from within the energy efficiency community unanimously agree that net savings are those that would not have occurred without the program. Yet the operational definition of net savings differs among programs and jurisdictions.

The literature and interviewees focused on two components of net savings: free ridership and spillover. Jurisdictions differ in their approaches to these two components of net savings. Some allow for the inclusion of both in net savings, while others allow only free ridership to be counted. Also, some jurisdictions expect free ridership and spillover to be isolated and measured, but others allow free ridership and spillover to be embedded in the estimate (such as for market-level estimates).

A related issue involved whether to consider adjustments such as installation rates, realization rates, leakage, rebound, and hours of use as part of net savings or as adjusted gross savings to which a net-to-gross ratio would still be applied to yield net savings. For the purposes of this paper, the authors treat net savings as referring to the adjustments associated with free ridership and spillover only, while the other adjustments are treated as components of adjusted gross savings. The recommendations section touches on this issue.

---

conference paper, but refer interested readers to the full paper (NMR & RIA 2010) as well as to Friedmann 2007; Hoefgen 2010; Messenger *et al*. 2010; Peters & McRae 2008; Saxonis 2007; Skumatz, Khawaja, & Colby 2009; and TecMarket Works 2004, among many other thoughtful articles and books that have been written on the subject of net savings and free ridership.

Table 1. Audiences for Net Savings Estimates

Audience	Established or Emerging	Use of Net Savings Estimates
Energy efficiency program administrators and planners; energy regulators; legislators; advocacy groups	Established (in some areas legislators and advocacy groups are emerging audiences)	• Assess if program achieved savings goals • Identify strong and weak areas of program design and redesign program accordingly • Apply strong program designs for other products, in other jurisdictions • Adjust payments to / funding of programs based on goal achievement • Determine if the ratepayer / taxpayer funds are being spent cost effectively and wisely
Air regulators	Emerging	• Will apply emissions factors to energy savings to estimate greenhouse gas reductions • Assess degree to which efficiency programs have achieved greenhouse gas reduction targets • Disagreement over whether will require adjusted gross or net savings

Estimation of the Counterfactual: The estimation of net savings typically relies on quantifying what would have happened in the absence of the program, which is the third contextual issue. However, because the program did happen, evaluators must estimate this counterfactual. 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.

Some methods, however, likely produce more valid estimates of the counterfactual, and, therefore, net savings than others do. For example, analysis of longitudinal sales data in areas with and without programs offers a promising approach for isolating net savings for programs expected to have market-level effects (*e.g.* ENERGY STAR® appliances and lighting programs). Unfortunately, manufactures and retailers—including program partners and recipients of the ENERGY STAR Partner of the Year award—often refuse to provide sales data that allows for a careful assessment of how program activity has influenced sales of the technology. Without such data, evaluators regularly fall back on self-report based approaches, which many commentators believe are more susceptible to measurement error, thereby producing potentially biased results that suffer from a lack of validity. While data and methodological limitations pose problems for all programs, the nature and severity of the problems differ with the type of program. Measurement of spillover, for example, may be more challenging for mass market measures.

Net Savings in Relation to Current and Evolving Policy Needs

Recent legislation and policies targeting climate change, national energy independence, and economic stimulus have combined with traditional energy efficiency program drivers to bring about expanded goals for both energy savings and non-energy benefits, such as greenhouse gas reductions and job creation. Moreover, individuals are being exposed to an increasing number of programs and

messages encouraging efficiency, while national economic trends also have an effect on actions related to efficiency. This situation increases the challenges of attribution. This section discusses commentators' views on current and evolving policy needs as well as the implications for determining attribution.

Meeting Current Policy Needs. Commentators had mixed opinions about the extent to which existing net savings approaches meet current policy needs, with energy regulators being more likely to say that existing approaches meet current policy needs, and program administrators and net savings experts voicing greater skepticism. Those who thought that current approaches are sufficient to meet current policy needs tended to focus on the importance of measuring net savings estimates. They argued that net savings estimates and insights gained from process evaluations and free-ridership and spillover studies allow program administrators, planners, and regulators to understand the strengths and weaknesses of a program design and decide whether and when to revise or discontinue a program. Further, these commentators said that net savings estimates ensure that ratepayer and taxpayer funds are spent wisely so as to achieve the greatest return on investment in terms of energy savings. Commentators who were skeptical about the ability of current net savings approaches to meet current policy needs had two primary concerns. One was the contention that most of the measurement approaches currently in use could not estimate net savings at the level of accuracy and precision needed to meet current policy needs. The second concern was that net savings is too narrow a focus and fails to capture important aspects of programs, such as behavioral change, market transformation, and how well they engage the customer. A related concern was that the current focus on net savings—particularly free ridership—could inhibit innovation in program design that will allow programs to meet evolving policy needs.

Meeting Evolving Policy Needs. The authors also 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. Their responses 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) No one is sure what these changes will entail or what will be needed, so the direction to take remains uncertain.

This uncertainty is perhaps most notable regarding air regulation. There is the potential for—and some commentators said a high likelihood of—a federal greenhouse gas regulatory program to exist in the future. Multiple commentators argued that the energy efficiency community should plan for this eventuality by working with the air regulation community to develop evaluation and reporting approaches that support both communities. 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 the air regulation and energy efficiency communities about the degree to which current energy efficiency evaluation practices do or can meet air regulatory needs. A pivotal issue is whether the air regulation community will be required to base savings on adjusted gross or net savings; if net savings are required, the air regulation community will also have to address the challenge of attribution of savings and of emissions reductions. A related concern is which methods of estimating energy savings will be acceptable for "proving" that emissions reductions actually occurred. Another issue is how to define the baseline in projections of emissions, particularly whether those baseline projections start out by assuming the emissions reductions resulting from energy efficiency programs or if such reductions will be credited to the programs as achievements beyond the baseline.

Attribution in the Face of Multiple Programs and Policies. Increasingly, the energy efficiency community is focusing a great deal of attention on attribution: how to attribute program impacts to particular programs when there are multiple additional factors influencing the behaviors targeted by the program. 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.

One of the primary challenges of attribution is that people are not necessarily aware of the causes of their own behavior and their explanations can therefore be inaccurate. For example, after taking an action, individuals tend to see themselves as the "sort of person" who takes that sort of action, minimizing the influence of external factors, such as a program's incentive, that in fact might have influenced the action in the first place. Or they might have intended to take the energy-saving action promoted by a program before hearing about the program, but might not have been motivated actually to take the action until exposed to the program or its incentive. The question in this example becomes how much credit to give to the program versus the other influences that had existed but had not motivated them to action.

Some commentators stressed that, although it is important to disentangle the effects of a particular program from other influences, it is also 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. These sources also warned 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.

Consistency in Net Savings Definitions and Methods

When sponsoring this effort, the Forum was particularly interested in the question of whether or not promoting consistency in net savings definitions and methods across the Northeast was a task it should take on, and, if so, how the Forum might go about pursuing consistency. The project team explored opinions about the possibility of adopting consistent definitions of and methods for measuring net savings throughout the Northeast.

Most commentators supported the idea of having a consistent definition of net savings in the Northeast (*i.e.,* which components are included). The main reason the region currently does not have a consistent definition is that there has been no explicit public policy driver, such as legislation, to stimulate its development. Instead, jurisdictions have simply developed their own definitions, and may be reluctant to let go of them without some compelling reason to do so.

Opinions diverged, however, on whether methods for measuring net savings should be consistent, given a particular definition of net savings. Most respondents recognized some benefits of adopting consistent methods, particularly in light of increased regional cooperation on 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 could stifle creativity and may not reflect the diverse range of programs and varying resources. 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.

In general, air regulators were the strongest advocates for consistency in estimation methods. Net savings experts and program administrators generally supported consistent methods for net savings, but raised concerns about the reliability and validity of current 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, pointing out that methods should keep evolving and improving before any particular methods are prescribed.

Conclusions and Recommendations to the Forum

The research presented in this paper demonstrates that the expanding audiences for net savings estimates and the evolving policies that may influence the measurement and use of net savings will present new challenges to the already confounding problem of net savings estimation in the energy efficiency community. The research presented here and elaborated on in the full scoping paper supports the continued use of net savings estimates for four specific purposes.

The first is a*ssessing the degree to which programs cause a reduction in energy usage and demand*. 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 savings or adjusted gross savings that recognizes that some energy and demand savings would have happened without the program.

The second use 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, 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.

The third involves *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.

The final continued use supported by this research is to *understand better how the market responds to the program* and to use 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. The authors believe that these recommendations, while made to the Forum, have broader application in the energy efficiency and air regulation communities.

Recommendation 1: Lead the process of developing consistent definitions of and approaches for measuring adjusted gross savings and net savings in the Northeast Region. The research uncovered a wide variety of both definitions and conceptualizations of net savings and its components as well as approaches to estimating net savings. The majority of sources considered this lack of consistency to be problematic and supported the long-term goal of achieving more consistency across the Northeast in how net savings is defined and measured. Because of the consensus that consistency is unlikely to be achieved without a policy driver, the authors recommended that the Forum lead the way toward greater consistency. The first step in achieving the goal of more consistent definitions of net savings is to define the elements of adjusted gross savings, to which net savings are applied. The next step is to define the elements and concepts of net savings, allowing everyone to "speak the same language," but not necessarily requiring that everyone measure every element. Research to support this recommendation would focus on understanding current reporting needs and practices for ISOs and forward capacity

markets, regulators, and program administrators. The next step would be to enumerate similarities and differences in definitions and requirements across the region and to develop consistent definitions that would meet the needs of the diverse energy efficiency community in the region.

Recommendation 2: The Forum and its allies should consider taking action to improve the quality of data used to estimate net savings. This may involve 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 at the county or state level. It may also involve encouraging program administrators to keep records of program activity by year, including in any possible comparison areas. The research revealed a great deal of dissatisfaction with net savings methods relying on self-reported free ridership and spillover or surveys of comparison areas. Several sources noted that market-based approaches using sales data avoid these problems by embedding free ridership and spillover into the estimates. However, due in part to confidentiality concerns, inconsistent access to the types of sales data needed for accurate estimation of net savings severely limits the use of such methods. Legally requiring reporting of these sales 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 approaches.[4] Note that this recommendation does not require further research, but, instead would focus on advocacy and working to overcome the barriers to data access caused by confidentiality concerns.

Commentators also mentioned the challenges of gaining access to longitudinal data, including program data (*e.g.,* budgets, measure covered, participants served, and previous estimates of net savings, among others) that could assist in the estimation of net savings and cumulative program. More careful record keeping of 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.

Recommendation 3: The Forum should clarify the definition of *attribution* and the degree to which programs must differentiate the impact of their activities from the impacts of other factors that may lead to the same or similar actions or outcomes. 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. Deciding the degree to which programs must (or can) disentangle their impacts from those of other programs and external influences will involve 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. Related research would include understanding the impact of participants who take part in multiple programs, including any possible synergies from participation in multiple programs, as well as identifying the variety of potential influences outside of program activity that may affect program participation with an eye toward establishing whether causal links can be drawn

[4] Examples include structured expert judgment (Delphi panels), weight-of-evidence, and historical tracing approaches.

between these influences and participation.

Recommendation 4: The Forum should encourage the energy efficiency community—particularly energy regulators—to expand its assessment of program success from a focus on net savings 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 the full scoping paper 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 the authors recommended the continued consideration of net savings estimates, they also suggested that additional impacts be taken into account when assessing programs and determining any reward or penalties to be paid. Research involved in expanding the focus of evaluation beyond net savings would include identifying the impacts other than energy savings that are most crucial to determining program performance and ensuring that ratepayer and taxpayer funds are being used responsibly, and developing a prioritization scheme of these impacts.

Recommendation 5: 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 by developing a framework for how to achieve it. 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 the approaches that should be avoided except in limited circumstances.

This recommendation involves three different research needs. The first is to develop an understanding of 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 rely on the best net impact evaluation practices available. The second involves 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. The third is an examination of 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 6: 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. Some of the sources consulted in this research predicted that the near future will bring policies requiring the development of energy savings estimation methods that meet the needs of the air regulation community. They also said that the needs of the energy efficiency, air regulation, and system planning communities would best be met if these methods were developed in partnership. The tasks for this partnership would include examining possible approaches for translating energy efficiency impacts into measurable pollution reductions by time of day and year and exploring ways of measuring net energy savings that would meet the requirements for reliability and precision to which both communities must abide. The first research needed to support this recommendation is to examine the possible methods or approaches for translating energy efficiency impacts into measurable pollution reductions from power plants by time of day and time of year. The second type of research would involve exploring possible approaches to measuring *net energy savings* that would meet any requirements for reliability and precision.

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, the authors 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 the energy efficiency community needs a radical rethinking of how it goes 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 in the full scoping paper, the authors briefly discuss alternatives to measuring net savings that recognize that program activity does not necessarily yield all the measured gross savings[5]. This leads to the final recommendation.

Recommendation 7: 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 2, 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 other structured expert judgment approaches. Importantly, the approach itself, and not just the net-to-gross number, should be negotiated and agreed

[5] In addition, one outcome from the release of the Scoping Study was formation of a Policy Subcommittee, in which Forum members will explore policy issues, including implications of current or new approaches to net savings.

on beforehand; it is even possible—as with the recent decision in Arizona to credit up to one-third of the savings from utility codes and standards 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, influenced by other 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.

References

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

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

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.* Berkeley, Calif.: Ernest Orlando Lawrence Berkeley National Laboratory.

NMR Group, Inc. and Research Into Action, Inc. 2010. "Net Savings Scoping Paper." http://neep.org/uploads/EMV%20Forum/EMV%20Products/FINAL%20Net%20Savings%20Scoping%20Paper%2011-13-10.pdf. Lexington, MA: Northeast Energy Efficiency Partnerships.

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?" *In Proceedings of the ACEEE Summer Study on Energy Efficiency in Buildings,* 5-219 to 5-234. Washington, D.C.: American Council for an Energy-Efficient Economy.

Saxonis, Bill. 2007. "Free Ridership and Spillover: A Regulatory Dilemma." Paper presented at the Energy Program Evaluation Conference. Chicago, IL. August.

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

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.*" Berkeley, Calif.: California Institute for Energy and Environment.

TecMarket Works. 2004. *The California Evaluation Framework.* San Francisco, Calif.: California Public Utilities Commission.

This foregoing document was electronically filed with the Public Utilities

Commission of Ohio Docketing Information System on

1/28/2014 1:53:36 PM

in

Case No(s). 13-1027-EL-UNC

Summary: Reply Comments electronically filed by Mr. Steven T Nourse on behalf of Ohio Power Company