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 Section 4928.66, Revised Code.	**)** **)** **)** **)** **)**	**Case No. 12-665-EL-UNC**

COMMENTS OF OHIO POWER COMPANY TO THE PUBLIC UTILITIES COMMISSION OF OHIO'S OCTOBER 3, 2012 ENTRY

INTRODUCTION

On October 3, 2012, the Public Utilities Commission of Ohio ("Commission" or "PUCO") issued an entry establishing a comment period in order to assist the Commission in its review of Evergreen Economics' report ("Evergreen Report" or "Report") filed by Staff on behalf of Evergreen Economics ("State Wide Evaluator" or "SWE") in this docket on August 29, 2012.

Ohio Power Company ("AEP Ohio" or "Company") appreciates the Commission's request for electric distribution utility comments on this topic. As discussed in the comments, the Company has significant concerns with Evergreen Economics and the Evergreen Report relating to the apparent misunderstanding of the extent of the role granted by the orders/RFP to the consultant (Evergreen Economics) by the Commission. AEP Ohio's major concerns include: 1) the Report's emphasis on and reporting of free-ridership estimates, 2) the recommendation to review projects prior to filing, 3) the SWE's ability to approve documents prior to filing, 4) multiple references to use a document that has not yet been approved by the PUCO for use in Ohio (the

Technical Reference Manual or TRM), 5) recommendations to re-calculate filed savings in contradiction to the October 15, 2009 order in Case No. 09-512-GE-UNC, and 6) applying In Service Rates to the gross savings for CFL's. In addition, the Company would like to point out an error in the Evergreen Report when it states that AEP Ohio did not provide data for the *e3Smart* program in 2010.

AEP Ohio appreciates the ability to provide its input on the recommendations made in the Evergreen Report. AEP Ohio offers these comments and states that its views or concerns could change depending on the facts and circumstances in the future. AEP Ohio offers these comments as a resource for the Commission in an attempt to provide the Commission some specific and relevant input on the recommendations and issues contained in the Evergreen Report.

GENERAL COMMENTS

The Commission opened Case No. 09-512-GE-UNC to address the protocols for the measurement and verification of energy efficiency and peak demand response measures. After a four-month public process, the Commission issued an order on October 15, 2009, in Case No. 09-512-GE-UNC which developed five protocols for the utilities in Ohio to follow. Three of those protocols are at issue in the Report and include:

1. Gross Savings
2. Baseline Determination
3. Prospective Values

AEP Ohio also objects to the SWE's attempts to unilaterally require action by the utilities. On page (i) of the Report, the SWE lists the five objectives of the Commission

in originally awarding the SWE contract. Specifically, the order dated March 17, 2010 in Case No. 09-512-GE-UNC listed the following objectives in paragraph 6:

> RFP No. EE-lO-PES-01 sets forth the scope of work for the Evaluator, including, but not limited to, the following:
> (a) evaluating and validating the electric energy savings and peak demand reductions resulting from each approved electric utility program and mercantile customer activity;
> (b) determining program and portfolio cost-effectiveness; and
> (c) conducting some program process evaluations of energy efficiency programs.

The order directing Commission Staff to issue the aforementioned RFP, dated January 27, 2010, in Case No. 09-512-GE-UNC states that Ohio reports gross savings until the Commission decides otherwise at a point in the future. Page number 3-10 of the RFP, Section 3.2.2.(h) reads:

> If required by the Commission or Staff, calculating Net Energy and Demand Savings, using benchmark net-to-gross (NTGR) values and/or "bottom-up" NTGR analyses, considering full, partial and deferred free-riders, free-drivers and spillover using interview techniques.

The SWE presumes without basis that net-to-gross estimates should be applied now, without addressing the full range of issues listed in the RFP or following the due process anticipated by the Commission prior to entertaining such a dramatic change.

It is the Commission that should address such a change, and only prospectively after hearing from interested parties and considering the relevant factors. In the Report, the SWE reported on free ridership for selected 2010 utility programs. It would appear that the SWE either is not aware of these Commission orders or thinks that they somehow do not necessarily apply to the SWE. AEP Ohio requests that the Commission clarify its direction and intent for the SWE with respect to work that is "in-scope" or required by

Commission order and not to unilaterally address free ridership without exploring the breadth of the matter as the Commission has affirmed in two previous orders.

KEY FINDINGS: NET IMPACTS

At page (iii), the Evergreen Report discusses Net impacts and the activities of the Independent Evaluator to develop a free ridership battery of survey questions.

AEP Ohio considers the lengthy discussion of free ridership in the first 10 pages of the Report inappropriate because the utilities in Ohio, by Commission order in Case No. 09-512-GE-UNC, dated October 15, 2009, pages 4-6, and order dated January 27, 2010, page number 3-10 of the RFP, the Commission explicitly addressed the joint issues of free ridership and spillover/free drivers. It ordered that until expressly addressed by the Commission, the utilities will report gross savings. Further, the Commission's orders held that one aspect, such as free ridership, would not be addressed without the inclusion of all sides of the issues (page 6, paragraph 16, October 15, 2009 order). 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.

Contrary to the plain language and intent of Commission orders, however, the SWE is unilaterally addressing the issue, and furthermore, has only addressed one of the two components that would provide a net impact assessment of energy savings attributable to the utilities' programs. AEP Ohio does not believe that the SWE should unilaterally address the issue or attempt to require the utilities to use a set of survey questions in their internal evaluations of programs without involving all stakeholders in a collaborative process. AEP Ohio does not accept that a single paragraph in a report by

the SWE presuming the time is now is a substitute for a Commission decision after conducting a collaborative process.

Further, 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) 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.[1]

Wilson-Wright *et al* [2] (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

[1] 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

[2] 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.

devotes a significant number of dollars to both program specific and general marketing campaigns in support of energy efficiency. Mahone's 2011 paper[3] 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).

On page (iii), the SWE implies that the gross reporting requirement in the order specifically addressed reporting for 2009-2011. That is simply not the case and there was no expiration date for using the gross impacts approach. It is open-ended. AEP Ohio strongly disagrees with the SWE that the time has come to change the Commission's guidance to the utilities.

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 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, the Commission must determine that the time has come to re-visit the issues in Case No. 09-512-GE-UNC, not the SWE. Considerable time and money has been expended in evaluation research to determine the extent of free riders/free drivers in the past thirty years.

[3] 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.

Third, free riders and how to estimate it accurately continues to remain an open research question. The SWE discusses the difficulty of designing appropriate survey questions on pages 62-63 of the Report. 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.

Fourth, the Midwestern states do not have the same history of energy efficiency that exists in other parts of the country. In recognition of that fact, six states require only gross calculations, one uses both, and six of the states require net to gross calculations. Of those using net-to-gross estimates, five require estimates of spillover, which is consistent with Commission orders in Case No. 09-512-GE-UNC. The proposal in the Report does not address spillover, even though the Commission orders stated that the issues should be addressed together. Moreover, even when net-to-gross estimates are used, many times they are assumed or calculated to be 1.0.

Fifth, 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.

Sixth, 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.

The SWE addresses the scope of its objectives as defined by the RFP issued by the PUCO in 2010 on page (i) of the Report. Objective five lists several items of interest to the Commission that could be explored. Free ridership is not among them, yet the SWE improperly attempts to pursue it rather than any of the specific objectives listed by the Commission.

SCOPE OF STATEWIDE EVALUATOR'S ROLE TO DIRECT UTILITIES TO UNDERTAKE ACTION INDEPENDENT OF A COLLABORATIVE PROCESS

In fact, separate from the Report being discussed in these comments, the SWE has already sent the utilities a brief set of "required" questions relating to free ridership that it expects the utilities to use. On pages 4 and 9 of the Report, AEP Ohio believes that the SWE has exceeded its role. The SWE is directed to perform independent evaluations of utility programs. In the spirit of cooperation and to minimize the overall cost of evaluation and the number of times customers are contacted, AEP Ohio provided the SWE with draft copies of the evaluation plans and customer surveys for review and comment. At no time did AEP Ohio believe that it was submitting the surveys/plans *for approval* by the SWE as stated on page 4. AEP Ohio believed this information sharing was beneficial for all parties. The SWE improperly attempts to unilaterally direct the utilities to comply with evaluation plan formats, report formats, table formats, additional questions, etc., all without the benefit of a collaborative process involving all stakeholders.

Since there was no notice or process afforded to the utilities and such a change could only be made prospectively, AEP Ohio requests that the Commission declare that the SWE's purported directives on this point are simply invalid. AEP Ohio asks the Commission to reiterate for the SWE that its role is to perform independent evaluations and to review the Portfolio Status Report and the associated evaluations filed by AEP Ohio and not to direct the internal evaluations of the utility. AEP Ohio has cooperated with the SWE to limit customer intrusions and collaborate on surveys. However, the SWE still has conducted separate surveys.

On page 9 of the Report, the SWE states:

> In anticipation of future differences in interpretation for these types of projects, we propose that part of the Independent Evaluator role involve helping utilities and PUCO staff review the application savings calculations as they are being submitted for approval for those projects where there may be disagreement on determining the appropriate baseline. This would include mercantile customers and could also be extended to other large custom projects.

AEP Ohio does not object to the Commission staff using the SWE to assist in evaluating extremely large or unusual mercantile projects; however, it should be at AEP Ohio's discretion whether the SWE is consulted before mercantile projects are submitted to the Commission.

Also on page 9 of the Report, the SWE directs the utilities to use the draft Ohio Technical Reference Manual ("TRM"). Unilaterally directing the utilities to use a document that has not yet completed its regulatory review process is premature and inappropriate.

LACK OF CONSISTENCY IN INDEPENDENT EVALUATOR IN REPORTING ACROSS UTILITIES

Both AEP Ohio and Dayton Power and Light ("DP&L") are providing a program for K-12 students that is operated by the Ohio Energy Project. Each utility received the same program data from the implementation contractor and provided that information to the SWE. However, Table 6 at page 19 of the Report indicates that AEP Ohio did not provide participant information on this program while Table 11 at page 28 of the Report indicates that DP&L did provide participant information. AEP Ohio fails to understand how the same type of information from the same vendor can be considered sufficient for DP&L and insufficient for AEP Ohio. AEP Ohio asks that consistent standards be applied across utilities when the same implementation contractor and data collection standards are employed.

REPORT RECOMMENDATIONS FOR AEP OHIO

1. Apply a CFL installation rate to adjust savings.

Response: The Commission has determined that gross kWh savings are to be reported for compliance purposes. AEP Ohio views the installation rate adjustment as being inappropriate because it is: 1) a function of net-to-gross calculations, and 2) an accounting exercise because its internal research indicates that a large proportion of those not installed in year one will be installed in the future. Because energy savings are reported on an annual basis by AEP Ohio, including the expected energy savings in year one is appropriate. To hold over expected kWh to a future year to count for compliance will become an accounting nightmare. The approach of counting the energy savings in the year the measure was sold simplifies the counting issue. Mahone (2011) notes

(Attachment 3 at page 7) that, "(S)ome jurisdictions adopt a policy which requires savings to be counted only for measures that are recruited for the program, and are then completed, within a single program cycle. Even for a simple measure such as CFL replacements, recent analysis has shown that significant savings can come in later program cycles."

2. Ensure mercantile customer savings are calculated using the correct baseline assumptions.

Response: AEP Ohio certainly agrees with using the "correct" baseline assumptions to calculate energy savings. However, AEP Ohio has two objections to the recommendation by the SWE on page 24 that AEP Ohio recalculate the energy savings of a specific project:

1. Disagreement between the parties concerning the determination of "correct."
2. The recalculation is contradictory to the October 15, 2009 order in Case No. 09-512-GE-UNC.

First, considerable disagreement over what situation, retrofit or new construction, should be used as the appropriate baseline continues to exist between the SWE and AEP Ohio engineers, AEP Ohio's independent evaluators, the implementation contractor's engineers, and the engineering staff at the company where the measures were installed. On page 23 of the Report, the SWE notes that "the Independent Evaluator team accompanied the AEP Ohio evaluation team on a series of mercantile on-site visits in 2010." While both parties were present on-site, the conclusions derived by the SWE and

the AEP Ohio team were diametrically opposed. The SWE states in the Report, again on page 23:

> In practice, there is sometimes disagreement among evaluators on what the appropriate baseline should be – either replacement or retrofit – and the effect on savings can be significant, depending on which baseline is assumed. This issue arose with one mercantile project for AEP Ohio. In this case, the AEP Ohio evaluator assumed the existing equipment baseline, while the Independent Evaluator believes that a new equipment baseline is appropriate. In this instance, the different baseline assumptions resulted in a 74 percent decrease i[s](sic) savings estimated for this one very large project when the new equipment baseline was used.

The SWE contends on page (i) of the Report that their charge is to "develop independent savings estimates of program savings" while the RFP requested the respondents to determine energy savings "by independently calculating and validating the documentation provided by the utilities." AEP Ohio contends the documentation provided by the utility and during the site visit supports the retrofit baseline assumption.

Further, AEP Ohio contends that recommending recalculation ignores the Commission order of October 15, 2009 in Case No. 09-512-GE-UNC where the Commission held that (Page 11, paragraph 32), "for compliance purposes, and in order to provide certainty and predictability…ex ante estimates should be used…" Also, in paragraph 34, the Commission discusses updating the TRM to reflect evaluation findings. While the State of Ohio currently does not have an approved TRM, in the spirit of the order, AEP Ohio has updated its internal estimates of costs, energy, and demand savings based on the results of its evaluations.

Moreover, in a September 15, 2010 Entry ("September 15, 2010 Entry") in Case No. 10-834-EL-POR, the Commission adopted an eighteen-month pilot program that established a process for automatic approval of qualifying applications by mercantile

customers that commit energy efficiency, demand reduction, or demand response programs for integration with an electric utility's energy efficiency programs required by Section 4928.66, Revised Code. As part of the pilot program, the Commission approved an applicant's use of the "as-found" method for establishing the baseline for all energy efficiency calculations. In its October 31, 2012 Entry on Rehearing in Case No. 10-834-EL-POR, the Commission recently reaffirmed the as-found method. Under the as-found method, the baseline for energy savings is the efficiency rating of the existing equipment at the time of replacement. This approach requires the Commission to review the impact of considering equipment on an as-found basis upon the ability of the electric utilities to meet their benchmarks and upon the costs of compliance with the benchmarks. The recommended denial by the SWE of significant energy savings – found to be a retrofit application by AEP Ohio's evaluator Navigant (well respected in the industry) – is a prime example that ignoring as-found principles raises the costs of compliance with the benchmarks for the Commission to consider. Finally, when two qualified evaluators disagree as in this situation, AEP Ohio feels the customer and utility should receive the more favorable outcome. Attached is Navigant's response to the SWE's evaluation of the project (see Attachment 4).

The establishment of the pilot program nine months after the filing of the project at issue certainly reflects the evolution of the Commission's approach to the as-found baseline. In addition, the Commission accepted an application for re-hearing on the as-found baseline on December 9, 2009, in Case No. 08-888-EL-UNC (one day before the effective date of the rules in the Docket which rejected the as-found baseline) under

which the rules were established. The issue of the appropriate baseline for mercantile projects remained an open issue at the time AEP Ohio filed the project in question.

AEP Ohio recommends the Commission allow the full savings verified by Navigant and reported by AEP Ohio on the project in question.

3. Develop a complete list of sources for *ex ante* savings values.

Response: The measures included in the program are voluminous. A single business program may contain hundreds of eligible measures with hundreds more potential base cases. These estimates do not neatly point to a single source or a filed planning document such as AEP Ohio's *Volume 2: 2012-2014 Energy Efficiency/Peak Demand Reduction (EE/PDR) Action Plan*. The recommendation is unclear whether each individual savings estimate for each measure in the portfolio must be documented or a general reference to the category of measure for a given vendor will suffice.

For instance, in a business program, the same lighting measure will have different energy savings estimates depending on the building type. The variant in the energy savings for each building type is the operational hours of the business/measure and not the difference in the wattages of the pre-post conditions. AEP Ohio asserts that developing a source list of each *ex ante* savings value is burdensome. Unless a deemed savings value is used for the *ex ante* savings value (which AEP Ohio agrees should be sourced), AEP Ohio and its contractors use standard engineering equations to calculate energy savings. Rather, a general list of sources used in a program to estimate key assumptions is more appropriate.

4. Adopt the recommendations presented in the 2010 AEP Ohio evaluation report.

Response: The recommendation provided on August 29, 2012, 17 months after the initial recommendations were filed with the PUCO on March 15, 2011, is untimely and infringes on the Company's right to manage internal operations. AEP Ohio reviews the recommendations contained in the evaluation reports and reviews these with the evaluation and management teams for each program following the filing each year. The majority of these recommendations are implemented during the year AEP Ohio receives them or planned for future implementation when feasible. In the few cases where AEP Ohio and/or the implementation contractor may not implement a given recommendation, it is after a full discussion of the issues involved. AEP Ohio held the review of the evaluation reports submitted for 2011 program activity six weeks after the May 15, 2012 filing. However, if the report is adopted in its entirety by the Commission, AEP Ohio asserts such a recommendation violates the Company's right to internally manage its business.

CONCLUSION

Ohio Power Company respectfully offers the preceding comments to assist the Commission in its review of the Evergreen Report as filed in this docket on August 29, 2012.

//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 Ohio Power Company's foregoing **Comments of Ohio Power Company to the Public Utilities Commission of Ohio's October 3, 2012 Entry** was served by First-Class U.S. Mail upon counsel for all parties of record identified below this 2nd day of November, 2012.

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

Email Service List
William.Wright@puc.state.oh.us,
Ray.Strom@puc.state.oh.us,
etter@occ.state.oh.us,
Amy.spiller@duke-energy.com,
dboehm@bkllawfirm.com,
mkurtz@bkllawfirm.com,
jkyler@bkllawfirm.com
sam@mwncmh.com,
joliker@mwncmh.com,
fdarr@mwncmh.com,
mpritchard@mwncmh.com
Elizabeth.watts@duke-energy.com,
cdunn@firstenergycorp.com
stnourse@aep.com

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.

Free-Ridership as a Way to Kill Programs - How Evaluation Policies Can Frustrate Efficiency Goals

Douglas Mahone, Heschong Mahone Group, Inc., Sacramento, CA

ABSTRACT

As evaluators, we understand the importance of distinguishing net savings from gross savings. It is important that program dollars not be wasted paying for measures that would have been installed anyway. Those are the free-riders. In a simple world, programs are discouraged from, or even penalized for signing up free-riders. Policymakers may constrain their evaluators by adopting evaluation rules that narrowly define who is a free-rider and who is not. There are also directions from policymakers for program managers to seek market transformation, and so they are encouraged to influence customers and markets in many ways other than simple incentives: through training, public awareness, upstream market interventions, and multiple year relationships with customers. On top of that, there are influences from outside the programs. Evaluators understand that there are many influences on customers and their energy efficiency decisions. These make it challenging to parse out what efficiency actions would have happened absent the programs and what would not. This paper is a thought piece with examples of how evaluation policies can define free-ridership, with the result of killing valid energy efficiency efforts. The intent is to illustrate for policymakers the consequences of their evaluation policy rules, and to alert evaluators to ways that evaluation practices may have unintended consequences for the success of energy efficiency program portfolios. By adopting an overly broad definition of free ridership, policymakers and evaluators underestimate the savings from programs. This can lead to cut backs or cancellation of programs that are actually delivering savings that would not have occurred otherwise, thereby reducing progress towards ambitious goals (market transformation, zero net energy, reduced generation and other long term goals).

Introduction

Free-ridership is a concept with both good and bad consequences for energy efficiency. It has been the subject of many learned papers, years of debate among evaluators, and lengthy policy debates. The basic concept speaks to the prudent use of energy efficiency dollars: they should be spent to encourage customers to take energy efficiency actions that they would not otherwise take on their own. If program dollars are spent on people who would have taken the actions anyway, without program support, then those people are free-riders, and those dollars were misspent. Evaluators are tasked with studying this counterfactual, measuring how much of a program's resources were misspent on free-riders, and what the program savings were, net of free-riders. The consequences of free-ridership measurements vary. In the mildest cases, the information is used to refine program plans to better target customers and to assess progress toward market transformation. In the strongest cases, when free-ridership levels are deemed excessive, program managers are penalized, savings claims are discredited, and programs are cancelled.

These applications of free-ridership information represent legitimate policy choices. Policymakers need to ensure that efficiency dollars are well spent, and that they are not given to people who don't need them. However, policymakers need a sophisticated understanding of the limitations of free-ridership measurement, and they also need to understand how free-riders fit within the larger market context. Recognizing that some degree of free-ridership is unavoidable, and may indicate

progress toward market transformation, free-ridership is not always bad. Moreover, free-ridership is notoriously difficult to measure with precision because it is a measure of a counterfactual (what would have happened); it cannot be measured directly, and the indirect measurement methods are famously controversial among evaluators.

With free-ridership measurements, the devil is in the details. To ensure consistency in measurement of free-ridership, all evaluators (and some policymakers) set ground rules for how their evaluations will measure free-ridership. Depending on how conservative those ground rules are, the resulting measured free-ridership levels can be high or low. If very conservative, a high percentage of program participants will be found to be free-riders, and the converse is also true. If conservative measurement rules are adopted, however, the resulting high levels of free-ridership can come into conflict with other policy objectives. For example, it is increasingly common for policymakers to require aggressive program activities to achieve ambitious savings goals (e.g., the acquisition of all cost-effective energy efficiency, or saving 1.5% of system load). Over time, these activities reach a large share of the market and influence it through multiple program paths. In this context, overly conservative free-ridership measurement rules will find that much of the resulting savings are due to free-riders, with very low net savings. This may result in the premature discontinuation of energy efficiency programs and an increased likelihood that energy policy objectives will not be achieved. In effect, this approach will pre-maturely kill many promising program activities, and the ambitious savings goals will never be met.

Note, when we characterize free-ridership measurement rules as conservative or liberal, that we are not talking about precision of measurement. Precision is a function of evaluation methods and sample sizes. One can make precisely conservative measurements or precisely liberal measurements. We will show how net impacts can be biased downward by taking an overly conservative measurement approach.

The following sections provide examples of how free-ridership measurement rules can come into conflict with ambitious savings goals, as cautionary tales for those setting policies on free-ridership measurement rules. Due to the nature of this paper, we do not seek to name names or point to specific programs. The lessons we illustrate are applicable to a wide variety of circumstances in the current energy efficiency world. Also note that, as a matter of convenience, we will refer to utility programs when describing various kinds of efficiency programs; this is a shorthand way of referring to any type of energy efficiency program or market intervention, whether done by a utility, a third-party implementer, a non-utility portfolio manager or a government agency.

Ambitious Goals

Other papers have pointed out that energy efficiency goals are becoming increasingly ambitious. “A number of states have set savings goals for utility-sector energy efficiency in the range of 1.5% - 2.0% of total sales each year” (Kushler, York & Witte 2009). While few states are actually achieving these levels of savings, many are ramping up their capabilities and their program offerings. In doing so, it becomes clear that traditional, transaction based programs - pay an incentive for a customer to install a measure - are not likely to get to the broader and deeper savings required to meet the goals. It will require a variety of market interventions that can:

- Reach more market segments,
- Encourage more integrated measures,
- Increase customer awareness and knowledge,
- Engage customer organizations in multi-year efforts,
- Leverage market allies’, regional and national efforts,

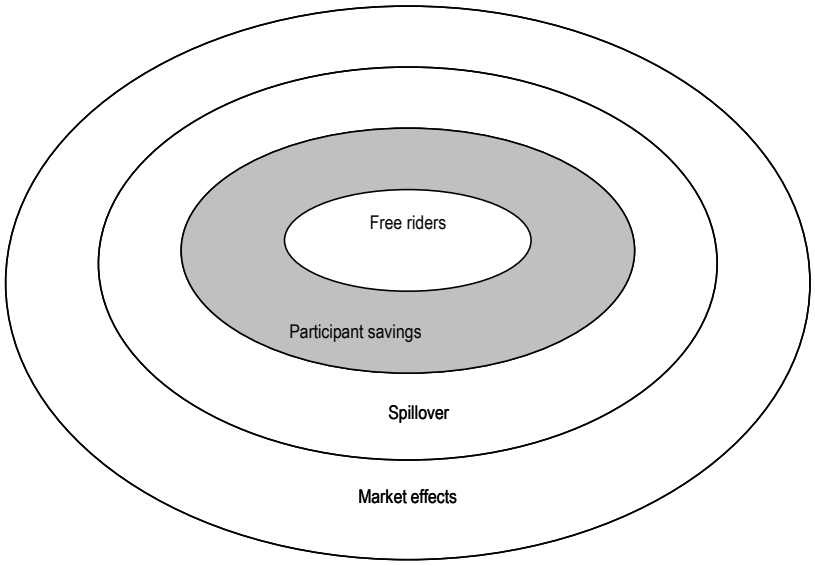
- Adopt more stringent codes and standards, and
- Provide public policy support to sustain those efforts.

This is not news to policymakers, program planners or evaluators. The challenge in meeting ambitious goals is to keep policies, program plans and evaluation procedures all pointed in the same direction. If there are conflicting directions, the entire energy efficiency effort will fail to meet the goals.

Interactions Between Programs

As we pointed out in our previous discussion of evaluation policy issues (Mahone& Hall 2009), the traditional evaluation paradigm assumes a measurement boundary around the program, as indicated in Figure 1 below.

Figure 1. Single Program Evaluation Components

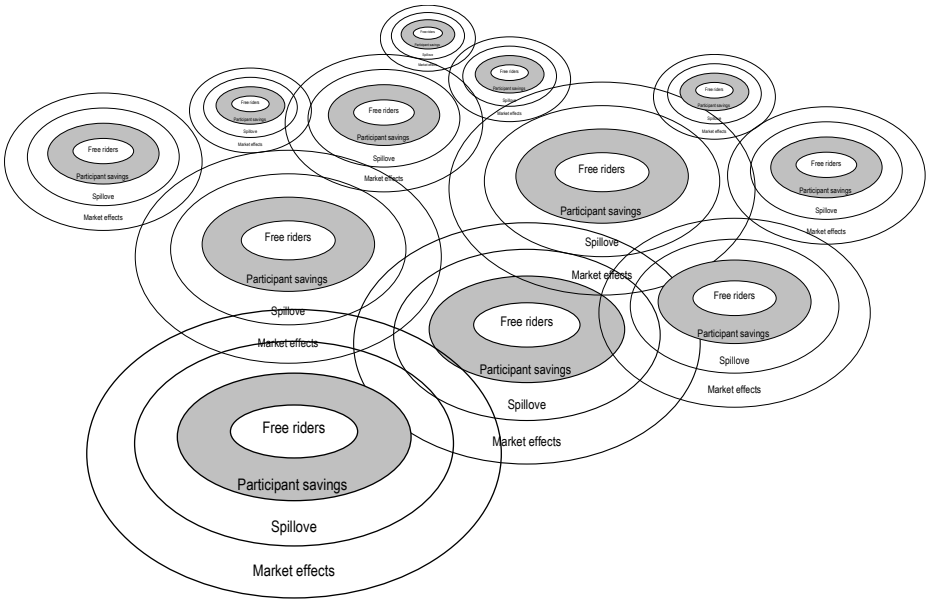


The participant savings, which are the verified savings from the efficiency measures installed through the program transaction with the customer, are shown in the shaded circle. Any of those savings attributable to free riders are subtracted out (inner circle), leaving the shaded donut of net participant savings. If there are spillover savings, either from additional measures installed by participants without program incentives, or from measures installed by non-participants who were indirectly influenced by participants' actions, these may be added to the total. Finally, as the programs begin to affect the market, e.g. by inducing retailers to sell only efficient equipment in response to market demand, there may be additional savings, called market effects. These additional savings would expand the shaded area of countable program savings.

This traditional approach to program design and evaluation worked well when efficiency program portfolios were limited in scope and focused on simple transactions with customers. However, with the evolution of integrated portfolios of programs targeted at broad and deep savings goals, the diagram now looks more like Figure 2. Note that the overlaps are, in practice, even greater than this simplified diagram would suggest. Often, the participant savings also overlap between program

offerings, and one program's participants may be another program's free-riders. The overlaps can also occur over time, with one program's influence extending to the next generation of programs. The point of the diagram is to suggest that the simple evaluation strategy of drawing a boundary around each program over a single program cycle will encounter the problems of multiple program influences. This can lead to biased estimates of savings, either higher or lower than actual.

Figure 2. Multiple Program Evaluation Overlaps



The following sections will illustrate various ways that free-ridership estimates can be biased by overly conservative measurement methods. When this happens, it can result in programs being penalized or even prematurely killed.

Killing Due to Unmeasured Spillover

Some methods of measuring free-ridership seek to compare efficiency measure adoptions among the participant population to a comparable non-participant population. For example, if 30% of the non-participant population is seen to be adopting a given measure without program support, then it is assumed that 30% of the participant population would also have adopted the measure on their own. These, then, become the free-riders in the participant population. This can be a problem if there is unmeasured spillover from the participant to the non-participant population. Continuing the example, let us say that half of the non-participants who adopted the measure (15% of that population in this example) did so because of indirect program influences. These could be influences from neighbors who participated in the program and were pleased with the measure, from customers who saw the program's information materials and were persuaded to adopt the measure but didn't want to hassle with the rebate paperwork, or from other program-related influences. If the free-ridership measurement does not account for these spillover influences on the non-participant population, it would, in this example, effectively double the level of free-ridership and overly penalize the program. It becomes a double penalty: the program is not credited for the spillover, and then it is penalized again because spillover is treated as free-ridership and further reduces program savings. While it can be difficult to measure the level of spillover, it is clearly more than zero, and it can significantly skew free-ridership measurements

that are based on this comparison group methodology. Most evaluators agree that failing to report participant spillover savings results in a downwardly biased estimate of overall true net savings.

A more progressive free-ridership measurement would credit the program for spillover in both participant and non-participant populations, or at least would control for non-participant spillover in the comparison group.

Killing Upstream Programs

Several kinds of programs that promote energy efficient consumer products (CFLs, refrigerators, washers/dryers, etc.) have made use of so-called upstream incentives, where incentive dollars are directed at the manufacturer or the regional distributor to buy down the cost of the product before it goes to the retailers. This not only reduces the product cost directly, but there are also indirect savings due to reduced retailer markup (typically a percentage of wholesale product cost). This approach often includes additional incentives to the retailer to assist them in promoting the product and stocking it. The result at least reduces the first cost of the product for the consumer, and at best provides a variety of strong motivators for the customer to buy the product. This can even include eliminating stocks of competing, less efficient products on the retailers' shelves. Free-ridership measurements are often done through customer interviews at point-of-sale, asking them what their awareness was of the products' benefits, which benefits (cost, convenience, performance, etc.) were most important, what influenced them most to make the purchase, etc. If the free-ridership measurement is looking for direct evidence that the program influences were the primary motivators for purchase, the program may be in trouble. With this kind of program, customers are often clueless about the program efforts and dollars spent upstream. Even if the program logo is affixed to the product or the product display, they may not notice it. Furthermore, the customer's awareness of the product benefits may have come from indirect utility sources (advertising, bill inserts, etc.), with little memory of where that information come from. If the free-ridership measurement gives considerable weight to that sort of awareness, the program is deeper in trouble. The effect could be severe enough to force program managers to discontinue their upstream incentives.

As we indicated in Figure 2 above, the multiple, overlapping influences on customers will make clear attribution difficult for these sorts of upstream programs, which could result in high free-ridership rates. A more progressive free-ridership measurement would not be limited to gauging customer memories of influence, and would make further measurement of the multiple program and non-program influences on consumer decision-making before assigning free-ridership numbers and reducing program net savings estimates.

Killing Organizational Efficiency Programs

There are programs that work with institutional or corporate customers to develop internal policies to maximize energy efficiency as equipment is replaced or new facilities are developed. In doing so, the customer organizations pre-determine that decisions will be made to seek extra levels of efficiency. These programs also encourage those customers to take advantage of utility offered incentives in making their efficiency investments. This all seems effective, even prudent, for programs as a way to work within the institutional decision-making structures of large organizations, particularly where a customer is expected to contribute toward the cost of the efficiency investment.

The problem arises when subsequent projects undertaken at the customer organization are evaluated for free-ridership. Often, the free-ridership measurement entails interviews with current staff at the organization, and often the staff say, in effect, "Of course we would have made those efficiency decisions. We were following established company policy to take advantage of incentives and make our

project as efficient as we could." A conservative approach to free-ridership would rule that the projects represented free-ridership, because the customer clearly would have done it under their existing policies. This, then, would downgrade program savings and could kill the program for having high levels of free-ridership. Even if the program was not killed, program managers would get a clear signal that it was a waste of program resources to encourage customers to adopt energy efficiency policies on the promise of future incentives, because they would be cancelling out their future savings; they would be deemed free-riders.

A more progressive free-ridership methodology would account for the utility role in the development of the energy efficiency policies, and would treat the subsequent efficiency decisions as the extension of those program activities. This would have the opposite effect to killing the program, and would encourage program managers to help overcome organizational barriers to energy efficiency.

Killing Through Complementary Government Policies

Another variation of an organizational efficiency program is a complementary government program. These are efficiency programs adopted by governmental organizations that seek, themselves, to leverage utility programs. One example is an affordable housing agency that requires all applicants for government tax subsidies to present projects that beat the energy code efficiency levels by some percent, fully intending that those applicants would also qualify for utility program support to achieve those efficiency levels. Another example would be a local government that adopts a resolution requiring all new buildings in their jurisdiction to show that they meet efficiency levels under a green rating system; again, with the intention that those buildings would qualify for utility program support. This should be seen as an admirable way to leverage influence on energy efficiency. The problem comes when these government agency policies become mandatory (or effectively mandatory); at that point the contribution of the utility program comes into question. The conservative judgment would be that, because participants must comply with the government agency's policy, then they become free-riders when taking advantage of utility program resources. Under that judgment, the local government program, even if carried out with the encouragement and support of the utility program, effectively kills utility programs for their participants, which may kill the savings despite the policy mandate. The policy question is whether the utility should be encouraging or resisting this kind of support from government agencies.

A more progressive free-ridership methodology would recognize the synergistic and mutually beneficial nature of these kinds of programs and policies, by at least sharing the savings success between the government agency and the utility. It could even be argued that the existence of the utility program and its resources is a necessary condition for the existence and success of the government policies, because those policies were predicated on the existence of program resources being available to ensure the cost-effectiveness of the policies they adopt. Experience with local government partnerships supports the claim that many governments could not mandate or enforce requirements for energy efficiency without the rebates, training, and other support provided by utility programs.

Killing Multi-Year Programs

Some of the biggest energy savings come from large, integrated projects which can take years to complete. A new hospital is a clear example, as are many comprehensive commercial retrofits. These projects achieve savings through interactive effects, economies of scale, and better integrated design than a series of individual measure installations could achieve. Even large residential new construction projects can take years to build out. These types of projects do not fall neatly within program cycles, even those that run for three or four years. For example, a three-year project would have to be initiated

in the first months of a three year program cycle in order to have any hope of completing within the cycle.

Some jurisdictions adopt a policy which requires savings to be counted only for measures that are recruited for the program, and are then completed, within a single program cycle. Even for a simple measure such as CFL replacements, recent analysis has shown that significant savings can come in later program cycles. For example, analysis of the 2009 upstream lighting program in California (KEMA 2011) estimated that 20 million CFLs rebated in the 2006-08 program cycle were installed in 2009. This represents over 20% of the total 2006-08 rebated CFLs. Ignoring the savings from these bulbs, because they do not fit neatly into one program cycle, unnecessarily reduces the program's effectiveness.

This approach makes even less sense for comprehensive projects such as major retrofits, which can take years to complete. Where such a policy is in place, a conservative approach to free-ridership would be to declare that any project completed in a later program cycle would not be counted in that later cycle, but would be deemed a free-rider, because it was responding to an earlier market influence. With such a policy, program managers would only focus on projects having a strong likelihood of completing within the program cycle. Large, long-timeframe projects would only be recruited in the opening months of a program cycle, and the program focus would shift to short-range projects as the end of the cycle approached. That's if the program wasn't killed outright for wasting resources on projects that couldn't produce measurable savings within the program cycle.

A more progressive free-ridership methodology would acknowledge prior program influences when they were clearly responsible for initiating a project, even if it has taken several years for the project to reach completion.

Killing Program Leveraging

Best practices studies and evaluator recommendations often urge program planners to leverage the education and promotional efforts of other organizations to help them reach and persuade customers to act on efficiency projects. For example, programs may ally with the federal ENERGY STAR® program, adopting its logo and its reputation in the market to help engage customers. Programs may enlist trade allies' training and marketing resources to complement their own. Programs may even adjust their marketing efforts to "sell the sizzle" instead of push the rebate transaction, emphasizing the non-energy benefits (comfort, improved indoor air quality, green bragging rights, etc.) that customers may value more than the energy or dollar savings. This alignment can be more effective at reaching customers, and so can represent a good use of program resources.

A conservative free-ridership measurement may seriously downgrade the net savings because of these program strategies. If customers report that they remember the ENERGY STAR logo or information more clearly than the program's marketing materials, that may be taken as an indicator that it was ENERGY STAR, and not the program, that was the motivator. The same could be true if customers remember the influences of trade allies more strongly than the program influences. Likewise, if customers report that they implemented the measure because they knew or cared more about its non-energy benefits, conservative free-ridership measurements may deem them to be free-riders, because it looks like they would have implemented the measure even without the program's rebate. These kinds of evaluation judgments can have the effect of discouraging program implementers from using leveraging strategies, as they end up being penalized for doing so. The result could be less cost-effective programs, fewer savings, or even dead programs.

A more progressive free-ridership methodology would credit the program for marshaling all those program leveraging influences to persuade the customer to participate in the program.

Killing Market Transformation

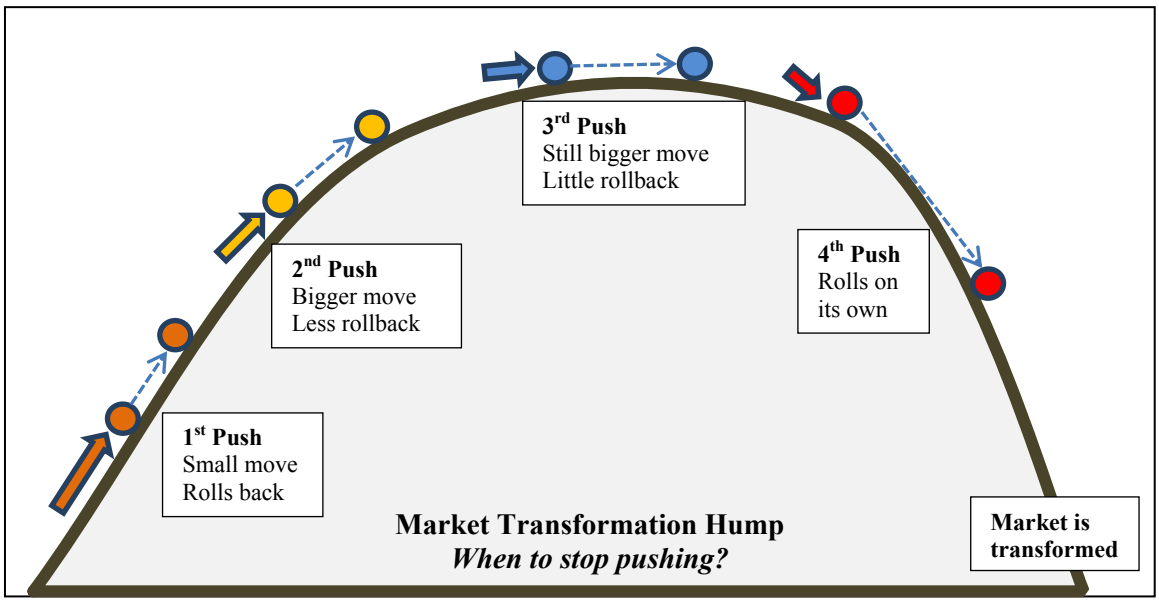
For most energy efficiency programs, the ultimate goal is market transformation, that magical point in the life of an energy efficiency measure when it has been embraced by the market and requires no further assistance from energy efficiency programs in order to continue with widespread adoption. When market transformation has been achieved, it becomes a waste of program dollars to continue pushing the measure. Virtually everybody can agree that this is clearly a time when no more program resources should be expended on the measure. The problem is in determining the point at which market transformation has been achieved. Over the lifetime of a product, and of the programs encouraging it, the rate of free-ridership will change as market conditions change.

Figure 3 provides a simplified way to think about this challenge, for a hypothetical efficiency product with a relatively short measure life compared to the life of a program that promotes it (i.e., the program will touch customers multiple times as they use the product). In the early days of an energy efficient measure's product life, there is a steep hump of market resistance to overcome. Customers are unfamiliar with the measure, contractors are not encouraging its use, suppliers are charging a premium, the measure is not widely available, etc. This is the point that an efficiency program may provide the first push to encourage market adoptions. That push may require a lot of education, high levels of incentives, demonstration projects, and other market interventions. It is also possible that a substantial fraction of the program participants are early adopters and free-riders; that is often part of the price of introducing new efficiency measures into the market. When that first push has run its course, if the program doesn't continue, the measure is likely to slip back in market adoptions and may get nowhere.

If the program continues with a second push on the measure, it will encounter less resistance and will make more progress in market adoptions. This is the point at which a conservative free-ridership measurement can do active damage. Part of the reason for the lesser resistance may well be that some of the influences from the first push are still resonating in the market, or that there has been some spillover from the first push. If the free-ridership measurements deem those first push influences on participants as causing free-ridership, then the success of the second push will be diminished, the program may be killed, and any progress in the market may be rolled back.

If, however, the program continues with a third push on the measure, it may encounter only small resistance and may win high levels of market adoption. Some might characterize the measure, at this point, as "low-hanging fruit" that the program should be de-emphasizing in favor of measures that are not as far up the adoption curve. If the free-ridership measurements find the influences on this third push to be largely hold-overs from the first and second pushes, and deems many of the participants to be free-riders, then, again, the program may be killed. This may not lead to a market roll back, but it may stall market progress.

Figure 3.Pushing Over the Market Transformation Hump



If the program survives to apply a fourth push on the measure, then it will find very little resistance, and will enjoy widespread adoption. If the consequences of the previous pushes are counted as free-rider influences, then most of the participants may be deemed free-riders, and the program would therefore be hit hard with very low net savings. The problem rests with the question of whether the product would ever get over the hump without that fourth push.

This example illustrates the difficulties of deciding when to pull the plug in the life of a successful program, and the role of free-ridership measurements in making that decision. If overly conservative free-ridership measurements are taken, and if free-ridership is used to penalize programs, then a lot of program efforts may be killed prematurely before market transformation or ambitious levels of savings are achieved.[1]

This example also illustrates that importance of estimating program spillover. As the measure is moved up the market transformation hump, spillover should increase as there are more satisfied customers, buying more of the measures even without the rebates, and influencing non-participants to do the same. Indeed, without the snowball effect of spillover, programs have a large burden to push the measure up and over the hump entirely on their own. While spillover can be difficult to measure, it is not zero, and it can be estimated.

Partial or Binary Free-Ridership?

These examples all beg the question of whether program-driven market interventions to encourage energy efficiency are the triggering factor in helping a customer to adopt a given measure at a given point in time. The basic concept of free-ridership is that, for those participants who are free-riding, something other than the program was the trigger, and their use of program resources was unnecessary or wasteful of ratepayer dollars.

[1] Of course, different products will have different lifecycles and different rates of adoption, so the trajectory of market adoption will vary. If this is accepted as normal, and the process of market adoption is considered by evaluators and policymakers, then programs and products can be smarter in adapting their approach to the market over time.

That said, few market observers would say that utility program interventions are ever the only factors needed to bring efficiency measure savings the market. To do so always requires other market actors: manufacturers, distributors, retailers, contractors, early adopters, organizational supporters, marketers, etc. Yet the simple, conservative concept of free-ridership implicitly asserts that efficiency programs should be the primary cause of action. If the programs get help from other sources, then their participants are free-riders to some degree. These are known as 'partial free-riders', and the extent of influence from other factors determines the degree of free-ridership. From this perspective, partial free-ridership measurements grossly under-estimate the degree to which these other market factors are needed to achieve program savings, and so they grossly under-estimate partial free-ridership. As a practical matter, every participant would be a partial free-rider.

An alternative way to measure free-ridership would seek to identify when the program influences were the trigger to achieving savings, rather than the degree to which they may have contributed. Under this view of free-ridership, program savings would be credited whenever the program's intervention was necessary for a given program participant to achieve the savings at that time. This would be a binary, yes/no determination for each participant in the program and the free-ridership study. The only free-riders would be the 'pure' free-riders who clearly would have implemented the measure on their own. A program's overall free-ridership would be based on the percentage of participants who were found to be free-riders. The concept of 'partial free-riders' would be abandoned.

Of course, the debate about free-ridership is much more involved than this very brief discussion may suggest. For a current summary of the issues, the reader is referred to (NMR &RiA 2010), as well as to (Friedman 2007), (Peters& McRae 2008), and others cited in that summary.

When to Kill Programs

None of this should be construed to say that programs, or the efficiency measures they promote, should never be killed. As markets transform, as products become mainstream, as customer behavior permanently changes, it becomes important to discontinue pushing for the old measures, and to refocus program efforts on advancing energy efficiency to the next level[2]. Every efficiency measure and program goes through a lifecycle, from initial rollout, through capacity building, through mainstreaming, to market transformation. As that process runs its course, free-ridership levels will change.

In the early stages, there may be high levels of free-ridership, because the only customers willing to adopt the measure may be the early adopters, many of whom would have installed the measures anyway. But by doing so, they are helping to build the market, to encourage suppliers to stock more efficient products, and to help installers become more familiar with the measures. This is part of the capacity building process.

In the mainstream stages, the programs can focus more on bringing in customers who could use the measures, but who are not using them for a variety of reasons that the program can help to overcome. Some free-ridership will still be present, inevitably, but if the program is bringing in too many free-riders it would indicate a need for targeting efforts to the other customers. This would probably be a bad time to kill the program outright, as it would throw away the investment made to get the market to this point.

Eventually, however, as the measures achieve widespread adoption, the levels of free-ridership are likely to increase again, unless program efforts can identify and target only the laggards who are slow to adopt efficiency measures. At some point it is time to pull the plug on the program and let the market continue without program dollars. Identifying the right time to pull the plug is not easy, but

[2] Of course, programs should also be killed for gross mismanagement or ineffectiveness. We're talking here about decently managed programs that are achieving savings and advancing the efficiency market.

good free-ridership measurements can provide important insights into customer behavior that will inform that judgment. If the plug is pulled too late, then program dollars are being misapplied. If the plug is pulled prematurely, use of the measure may stall or roll back, indicating that the market was not actually transformed and needed continuing support. After a premature program stop, going back into the market to re-start adoptions of the measures can be expensive, and in the meantime savings will be lost.

As this discussion illustrates, free-ridership can provide useful information on how and when to adjust programs and how they are delivered. It is not the only figure of merit for programs.

This all presumes a clear policy agreement about how to determine the right time to kill programs. Lacking such clarity, programs may die prematurely, for the wrong reasons. As the examples above illustrate, a program may be given poor grades for performance or cost effectiveness if the free-ridership measurement is based on overly-conservative criteria. The point is not that programs should never be discontinued, but rather that the criteria for doing so should be based on clearly articulated and applied measures of performance that do not inadvertently conflict with broader policy goals, such as market transformation or deep energy savings. Setting the rules for free-ridership measurement, and avoiding these kinds of problemsis a critical part of setting those performance measures and policies.

Conclusion

Clearly, there needs to be a lot of informed judgment to make good decisions about when to discontinue or to re-focus program efforts. Free-ridership measurements can help to inform those judgments. The caution we urge is in making the policy choices about how to make and use those measurements. If made too conservatively, good programs will be killed prematurely and savings will be lost. If too many programs meet this fate, then ambitious savings targets and market transformation goals will not be achieved. Evaluators have a role in helping policymakers to understand the implications of free-ridership measurement rules, because a simplistic understanding of free-ridership can lead to these unintended consequences.

To summarize the recommendations discussed in this report:

- Take care when drawing program boundaries for purposes of evaluation that,you account for overlapping program influences. Consider expanding the boundary to encompass all important program influences.
- When measuring non-participant efficiency levels, be sure to account for spillover from participants and program activities so as not to over-estimate the non-participants' or baseline efficiencies.
- When measuring free-ridership in upstream programs, be sure to assess the multiple program influences on customers (direct and indirect), and don't rely just on their direct perceptions.
- When assessing efficiency actions by corporations or organizations which have adopted energy efficiency policies, do not discount the role of the utilities and their incentives in the creation of those policies in prior years.
- When evaluating efficiency actions that were influenced by governmental policies and programs, do not ignore the role of utilities and their incentives in helping to create the value derived from these complementary policies.

- Do not ignore or unduly discount the influences on customers of program efforts that span different program cycles. Today's apparent free-rider may represent spillover from prior years.
- Do not ignore or unduly discount the effects of program leveraging strategies that produce increased program participation.
- Do not overly penalize programs for free-ridership that is a normal part of progress toward market transformation.
- Be aware of the problems of assigning 'partial free-ridership', remembering that program influences will never be the only ones that enable savings. Focus instead on determining when program influences were the trigger that produced savings.
- Make sure that policy rules for measurement of free-ridership are consistent with overall savings objectives, and that they do not unreasonably discount real savings.
- Have clear policy direction for when to kill programs, based on program lifecycles not just free-ridership.

References

Friedmann, Rafael. 2007. "Maximizing Societal Uptake of Energy Efficiency in the NewMillennium: Time for Net-to-Gross to Get Out of the Way?" Chicago: Energy ProgramEvaluation Conference.

Kushler, Martin, D. York & P. Witte. 2009. Meeting Aggressive New State Goals for Utility-Sector Energy Efficiency: Examining Key Factors Associated with High Savings." *ACEEE Report No. U091.Washington, DC:* American Council for an Energy Efficient Economy,.

KEMA, Inc., "CFL Memo". 1/13/11. Available at: http://www.cpuc.ca.gov/NR/rdonlyres/21FA8895-A754-4A6F-BEE8-0F776D10D7D7/0/AppendixR_CFLMemo.pdf .

Mahone, D. & N. Hall.*2009.* "Pursuit of Aggressive Energy Savings Targets - Aligning Program Design and Evaluation". *Proceedings of the 2009 Summer Study on Energy and Efficiency, August 2009*.Washington, DC: American Council for an Energy Efficient Economy.

NMR Group Inc., and Research Into Action, Inc. (RiA). 2010. *Net Savings Scoping Paper.* Prepared for the Northwest Energy Efficiency Partnerships EM&V Forum. Available at http://neep.org/uploads/EMV%20Forum/EMV%20Products/FINAL%20Net%20Savings%20Scoping%20Paper%2011-13-10.pdf

Peters, Jane and Marjorie McRae. 2009. "Free-Ridership Measurement is Out of Sync With Program Logic…or, We've Got the Structure Built, but What's its Foundation?"*Proceedings of the 2009 Summer Study on Energy Efficiency in Buildings. August 2009.* Washington, DC: American Council for an Energy Efficient Economy.

1211 Main Street
Grinnell, IA 50112

641-236-3848

NAVIGANT

Memorandum

To: **Jon Williams, Linda Ecker – AEP Ohio**

CC: **Stu Slote, Randy Gunn, Ryan Del Balso – Navigant**

From: **Roger Hill – Navigant Consulting**

Date: **January 18, 2012**

Re: **Response to Self Direct Program *ex-post* savings - Project Review**

Navigant has reviewed the evaluation from the Ohio statewide evaluator (SWE), Evergreen Economics and their subcontractor Michaels Engineering, of the upgrade project submitted in the Self-Direct Program for 2009. The crux of their opinion to disallow savings from the upgrades is two-fold:

1. The lack of information regarding the as-found or existing equipment, and
2. Their contention that the plant was obsolete, therefore they believe the project should be classified as new construction with the baseline equivalent to new equipment.

While we appreciate the position of the SWE, Navigant disputes both of these contentions.

Response to Item 1

The SWE points to PUCO guidelines for baseline determination that require documentation of every piece of existing equipment in the baseline. This methodology presumes upgrades can be isolated for measurement and verification so that IPMVP[1] options A or B[2] can be employed. The PUCO guidelines were never intended for a project like this one. An upgrade of the magnitude of this project precludes employment of M&V options A and B, and thus the need for detailed equipment lists.

Furthermore, IPMVP option D – Calibrated Simulation cannot be applied in this case. Option C – Whole facility analysis with billing data or other sub-metering is the appropriate M&V method, and it requires only billing data and confirmation that the site activities do not change and that the comparison of pre- and post-upgrade data are valid. Navigant's

[1] International Performance Measurement and Verification Protocol, Efficiency Valuation Organization, 2007.

[2] Option A – Retrofit isolation: key parameter measurement isolation. Option B – Retrofit isolation: All parameter measurement

representative on-site did ascertain that there has been essentially no change in the purpose and use of the facility. Materials are staged for the production of equipment in a comparable number with the same scale of supporting activities. Thus the billing data is adequate for the pre-installation baseline and M&V.

Response to Item 2

The statewide evaluator asserts the plant was obsolete and had no useful life. The SWE concludes the baseline should not be "as-found" as included in *ex-ante* savings estimates, but rather the baseline should be new construction and, in effect, "best practice".

Navigant acknowledges "market changes" would have eventually forced the existing equipment at the plant to be phased-out for upgrading for new equipment. The market, however, does not set a specific date as to when such a change occurs and the plant becomes obsolete. The market is fluid, not static, which results in a period of time over which a process may slowly become obsolete while continuing to operate successfully in the marketplace.

This appears to be the case with this project and it is clear the existing system had some amount of remaining useful life (RUL) before the market would have rendered the plant obsolete. It is our assessment that this project was done in acknowledgment of the market changing, but not in reaction to the market having already completed this change. A quick online check shows it is still possible to get similar product from other competitors. The existing system likely had a RUL of a minimum of one year, with a possible RUL of several years, depending on the marketplace.

Further, there is no evidence that faltering equipment life required removing equipment at the time the project was initiated. On the contrary, the existing equipment continued to produce all of the product needed during the planning and transition periods, indicating remaining life in the existing equipment. Concurrent operation of the product lines indicates early retirement is the proper scenario, and the savings calculations should use the removed equipment as the baseline with adjustments for remaining life. Employing a dual baseline in this case results in using the existing product line as the baseline for a period of time equal to its RUL; beyond the RUL of the existing equipment, the baseline would shift to new industry standard equipment.

Navigant also disputes the contention that the equipment installed through the upgrade project constitutes "standard practice" – in effect the minimum efficiency of new product of this type. The SWE offers no documents supporting this assertion or comparing this plant to another of its type and vintage. The project file, on the other hand, contains a letter from the Energy Star Award for Sustained Excellence characterizing the upgrades as "successful best practices". Thus, even if this project were new construction, the SWE has chosen a baseline that is not supported by their research.

Navigant finds other points in the SWE conclusions in error. These include the following:

The SWE asserts that a code-level chiller is the baseline for cooling manufacturing processes when temperatures are above 50°F to deliver 50°F water for the process cooling. This baseline is inaccurate because a cooling tower is not capable of delivering 50°F water at all hours below 50°F outdoor air temperature. Furthermore, a code-compliant chiller was used by the SWE for the baseline for process cooling loads; even though the as-found manufacturing floor did not receive cooled air and it appears that comfort cooling for offices was provided by other equipment. Chillers used for manufacturing processes do not have

code requirements for efficiency, thus the baseline should be the as-found chiller for the process loads. Comfort cooling equipment must comply with energy codes, thus the portion of the as-found cooling load used for comfort cooling should use energy code compliant equipment as the baseline. If the SWE continues to base savings on presumed retrofit isolation, the appropriate baselines should be used for respective loads.

The SWE disallowed production compressed air and HVAC savings based on the "obsolete process" argument asserted above. Navigant counters that the compressed air and HVAC equipment that served the production floor was not obsolete as evidenced by the continuing operation of the facility and production lines. Compressed air and HVAC savings should reflect early retirement of the as-found equipment the same as the production improvements.

Navigant's interpretation of the baseline is "as-found" equipment retired early, and we agree with the program administrator that whole-building analysis is the most appropriate M&V option, supporting the original *ex-ante* savings estimates. Since the billing analysis did not include the post-installation period of the lighting retrofits, those savings were not included in the *ex-ante* estimates. Please note that the savings from the lighting project should be submitted later.

Randy Gunn	Roger Hill
Managing Director, Energy Practice	Managing Consultant, Energy Practice

This foregoing document was electronically filed with the Public Utilities

Commission of Ohio Docketing Information System on

11/2/2012 5:07:08 PM

in

Case No(s). 12-0665-EL-UNC

Summary: Comments of Ohio Power Company to the Public Utilities Commission of Ohio's October 3, 2012 Entry electronically filed by Mr. Steven T Nourse on behalf of Ohio Power Company