Programme Attribution Analysis: Use of Self-Reports and Triangulation

Daniel Violette, Navigant Consulting, Boulder CO, US
Laura Agapay-Read, Navigant Consulting, Boulder, CO US

Abstract

Deciding on how to develop and present information that will be most useful for decision makers regarding the cost-effectiveness of energy efficiency (EE) programmes is a challenge in every evaluation. There are many decisions that need to be made in assessing overall cost-effectiveness, which includes an assessment of energy savings that are attributable to the EE programme. Methods for assessing attribution have received considerable attention in the literature. The selection of methods to be used in a particular evaluation are based on the type of program(s) being evaluated, availability of data, and also on jurisdictional or country-specific views regarding what constitutes credible attribution analyses. Reviews of methods for assessing attribution include the use of surveys of programme participants and trade allies as one approach for assessing attributable programme accomplishments. The use of survey information gathered from participants, non-participants, and market trade allies has been used to develop hypotheses and estimate factors that are used in the assessment of attribution.

This paper outlines a rationale for estimating net savings factors, e.g., free ridership (FR) and spillover (SO), which can be used to adjust savings values that use business-as-usual or common-practice baselines. A survey-based application is presented using three sources of information: (1) programme participant fast feedback surveys (conducted soon after the customer's decision to participate), (2) programme participant end-of-year telephone surveys, and (3) programme participant trade ally telephone interviews. This allowed for triangulation of results, tests of consistency, and sensitivity analyses – all part of a "best-practices" application. In the conclusions, several issues brought up in the literature are addressed concerning the reliability and cost-effectiveness of survey approaches for addressing attribution compared to other methods used to assess attribution.

Introduction

Approaches for estimating overall energy savings and attributable savings (also termed additional or net savings) have many common elements across Europe and North America. Almost every approach used in Europe is also used in North America; however, there are distinct differences in the philosophy and emphasis in the methods applied. There is general agreement about additionality as a concept, i.e., the energy savings that would not have occurred without the EE programme. However, technical issues in the implementation and estimation of additionality poses challenges.

A workshop on Article 7 of the Energy Efficiency Directive (See Austrian Energy Agency, 2015) held in Brussels addressed technical issues relating to additionality. Separate working groups were set up to address two issues: 1) Materiality &Additionality, and 2) Free Riders (FR). One finding was that definitions of materiality and additionality, as well as implementation of these attributed in a policy setting, is quite different among Member States. The working group discussion "showed clearly the lack of homogeneity regarding the concepts of additionality and materiality," and several participants underlined the need of having a common definition of both terms as a starting point. Similar concerns have been expressed in North America. Northeast Energy Efficiency Partnership[1] (NEEP, 2010) found that different regions and jurisdictions conceptualize "net energy

[1] The Northeast Energy Efficiency Partnerships (NEEP) facilitates the Regional Evaluation, Measurement, and Verification Forum (EM&V Forum or Forum). The Forum's purpose is to provide a framework for the development and

savings" differently, particularly with respect to free ridership and spillover. A corollary was that different programme administrators and evaluators relied on numerous methodologies to estimate savings making it difficult to compare findings, and even more difficult to assess the attainment of regional goals that span several jurisdictions. NEEP (2016) developed a Principles and Guidance document which emerged out of recognition among EM&V Forum member states of the importance of understanding how states define, estimate, and apply gross and net savings across the region

Several papers on attribution (or net savings) estimation methods have been produced.[2] These outline nearly a dozen different estimation methods, with two methods being the use of Business-as-Usual (BAU) baselines (also called Common Practice or Standard Practice Baselines), and the use of survey methods. The Brussels Workshop on the Energy Efficiency Directive discussed methods to evaluate additionality including: (1) business-as-usual (BAU) scenarios, and (2) surveys/questionnaires to consumers and businesses. These two methods are often considered both in the EU and North America.[3] The workshop suggested that a BAU scenario approach may be more reliable and practical than survey methods.[4,5] The right approach to use with a given programme, and within a given context and application always requires judgment and the consideration of the costs of the evaluation effort compared to the additional value of the information produced.

A discussion of the practical and cost advantages of BAU methods versus survey methods is presented in the conclusions section and after the presentation of the survey application. However, it is useful to consider some underlying assumptions of the BAU method versus survey methods that can directly address specific issues in additionality, notably free riders (FR). NEEP (2014) defines BAU as being based on the on the efficiency of current equipment commonly purchased in the market. Slote et al. (2014) defines BAU baselines "according to data on average energy performance of the eligible technologies and their level of market penetration." Equipment installations with energy efficiency that exceeds the BAU average energy efficiency is determined to produce additional/attributable energy savings.

One concern with the use of BAU baselines is the fact that they represent the average equipment purchase in terms of energy performance for that measure or end-use. An issue that often arises in the use of the BAU approach is termed self-selection bias. Self-selection implies, in general, that when customers are given the choice to participate in a programme, those that do choose to participate (i.e., select themselves into the program) may be systematically different than those that choose not to participate. One type of self-selection bias that may occur is based on an assumption is that those that choose to participate in an energy efficiency programme are those same energy users that likely would have selected higher efficiency equipment even if the programme had not been offered. If true, there would be higher FR and lower attributable savings than would be captured with a BAU approach. In other words, the assumption of the average customer being the right baseline for participants that select themselves into an EE programme may not always be accurate. The emphasis that evaluators have given to self-selection in North America has made them wary of the BAU and common practice baseline approaches. There is a belief that the market BAU may not represent the typical participant that selects

---

use of common and/or consistent protocols to measure, verify, track, and report energy efficiency and other demand resource savings, costs, and emission impacts to support the role and credibility of these resources in current and emerging energy and environmental policies and markets in the Northeast, New York, and the Mid-Atlantic region.

[2] These include: Violette and Rathbun (2014), GHG Institute (2012), Johnson (2014), Ecofys (2012), and Slote et al. (2014) among others.

[3] There is a growing literature on statistical, econometric, and experimental approaches to assess additionality, but these are generally limited to EE programmes with large numbers of relatively homogeneous participants such as residential customers (See Violette and Rathbun, 2014).

[4] See Austrian Energy Agency (2015, p. 24).

[5] Wuppertal (2009) states that typical method for estimating additional energy savings and free rider effects include "surveys of participants (and control group and other market actors) to find out reasons for implementing end-use actions." (See Report: "Measuring and reporting energy savings for the Energy Services Directive – how it can be done." Results and recommendations for the EMEEES project.)

themselves into an energy efficiency programme. Another issue is that there is How to assess any spillover[6] (SO) that may occur due to the programme using a BAU approach.

One way to address this concern is to estimate FR (or SO) directly to see whether participants that select themselves into a programme are more disposed to taking energy efficiency actions than the average customer, if the programme had not been offered. One way to accomplish this is to talk with customers, equipment suppliers and trade allies that specify equipment for different end-uses. This is typically accomplished through survey methods. The next section presents the application of a survey approach for the estimation of FR in a commercial sector EE programme at a mid-western utility in the United States.

Survey-Based Attribution Analysis – Commercial Sector EE Programme

The reported findings of the Workshop on Article 7 of the Energy Efficiency Directive (See Austrian Energy Agency, 2015) held in Brussels indicated that there was some concern about the reliability of survey analyses for producing estimates of FR. There is no question that there are concerns regarding the use of surveys to provide insights into additionality of savings. However, assumptions are made in any analyses of additionality. As shown above, there are some strong assumptions embedded in the use of BAU baselines to assess additionality. It may be determined that the use of the BAU concept is the method preferred in certain evaluation applications. However, it still might be wise to test certain underlying assumptions in the BAU approach regarding the baseline and its treatment of FR and SO. A survey of a sample of participants might provide useful information on the potential magnitudes of FR and SO not accounted for in the BAU approach, even if it is not used as the primary method for assessing attribution.

The benefits of a survey-based approach include:

- A survey approach can be less expensive than other approaches, particularly if the effort is combined with data collection activities that are already planned for process and impact evaluations.
- The evaluator has the flexibility to tailor questions based on variations in programme design or implementation methods.
- It can yield estimates of free ridership and spillover without the need for a nonparticipant control group (NMR Group, Inc. and Research Into Action 2010).

Despite these benefits and the wide use of a survey-based approach, concerns have been raised (Ridge et al. 2009; Peters and McRae 2008) including:

- A potential bias related to respondents giving socially desirable answers.[7]
- The inability of consumers to know what they would have done in a hypothetical alternative situation, especially in current programme designs that use multiple methods to influence behavior.
- The tendency of respondents to rationalize past decisions.

[6] Spillover refers to additional reductions in energy consumption or demand that are due to programme influences beyond those directly associated with programme participation. These spillover reductions may result from participants taking actions that are not directly rebated or included as a programme measure. This is termed participant spillover (PSO). There is also nonparticipant spillover (NPSO) where customers see that energy efficiency actions that participants are taking and decide to take similar actions, but do not officially participate in the programme. An example might be new construction where contractors see the best practices being used in a new construction programme and decide with their customer to implement some of these actions, but not directly participate in the program. This spillover effect has been widely documented in North American new construction and major remodel programs. As a result, these savings may not be recorded in the programme tracking system and credited to the program.

[7] Participants may also have a bias toward overstating programme impacts because they want to retain incentives, although this has not been widely documented.

- A potential for arbitrariness in the scoring methods that translate responses into free rider estimates.
- Consumers may fail to recognize the influence of the programme on other parties who influenced their decisions. For example, a programme having market effects may have influenced contractor practices, which in turn may have indirectly impacted the participants' (and nonparticipants') decisions.

Ridge et al. (2009) point out that, although these concerns are valid, they are widely acknowledged by social scientists who have worked on a variety of methods over the years to address them. It is also important to recognize that all methods have potential biases.[8] For example, market sales analysis[9] based on objective sales data can be biased if the market actors who provide data for the analysis operate differently from those not participating in the study or if the comparison area is systematically non-comparable.

It does not make sense to compare all survey self-report approaches equally, as some conform to best practice and others do not. Keating (2009) adds that many of the criticisms of the survey self-report approach can be alleviated through careful research design, sampling, survey timing, and wording of questions. The literature also contains a number of best practice elements for survey design, data collection, and analytic methods specific to estimating net savings (New York State Department of Public Service 2013; Tetra Tech et al. 2011).

Elements of a best practice survey approach can include (from Violette and Rathbun, 2014):

- Use questions that rule out rival hypotheses for installing the efficient equipment.
- Test the questions for validity and reliability.
- Use consistency checks when conducting the survey to immediately clarify inconsistent responses.
- Use measure-specific questions to improve the respondent's ability to provide concrete answers, and recognize that respondents may have different motivations for installing different measures.
- Use questions that capture partial efficiency improvements (accounting for savings above baseline but less than programme eligible), quantity purchased, and timing of the purchase (where applicable for a measure) to estimate partial free ridership.
- Use neutral language that does not lead the respondent to an expected answer.
- Establish *a priori* rules for treatment of missing/don't knows in the scoring algorithm.
- Weight the estimates by annual savings to account for the size of the savings impacts for each consumer.
- Sample, calculate, and report the precision[10] of the estimate for the design element of interest (measure, project type, or end use).
- Conduct sensitivity testing of the scoring algorithm.
- Define what the spillover measurement is and is not attempting to estimate and justify the use of an approach.
- Employ, where feasible, a preponderance of evidence (or triangulation of results) approach that uses data from multiple sources. Potential data sources could include project file reviews, programme staff and account manager interviews, vendor interviews, and observations from site visits.

---

[8] This is, of course, the primary motivation for triangulation.

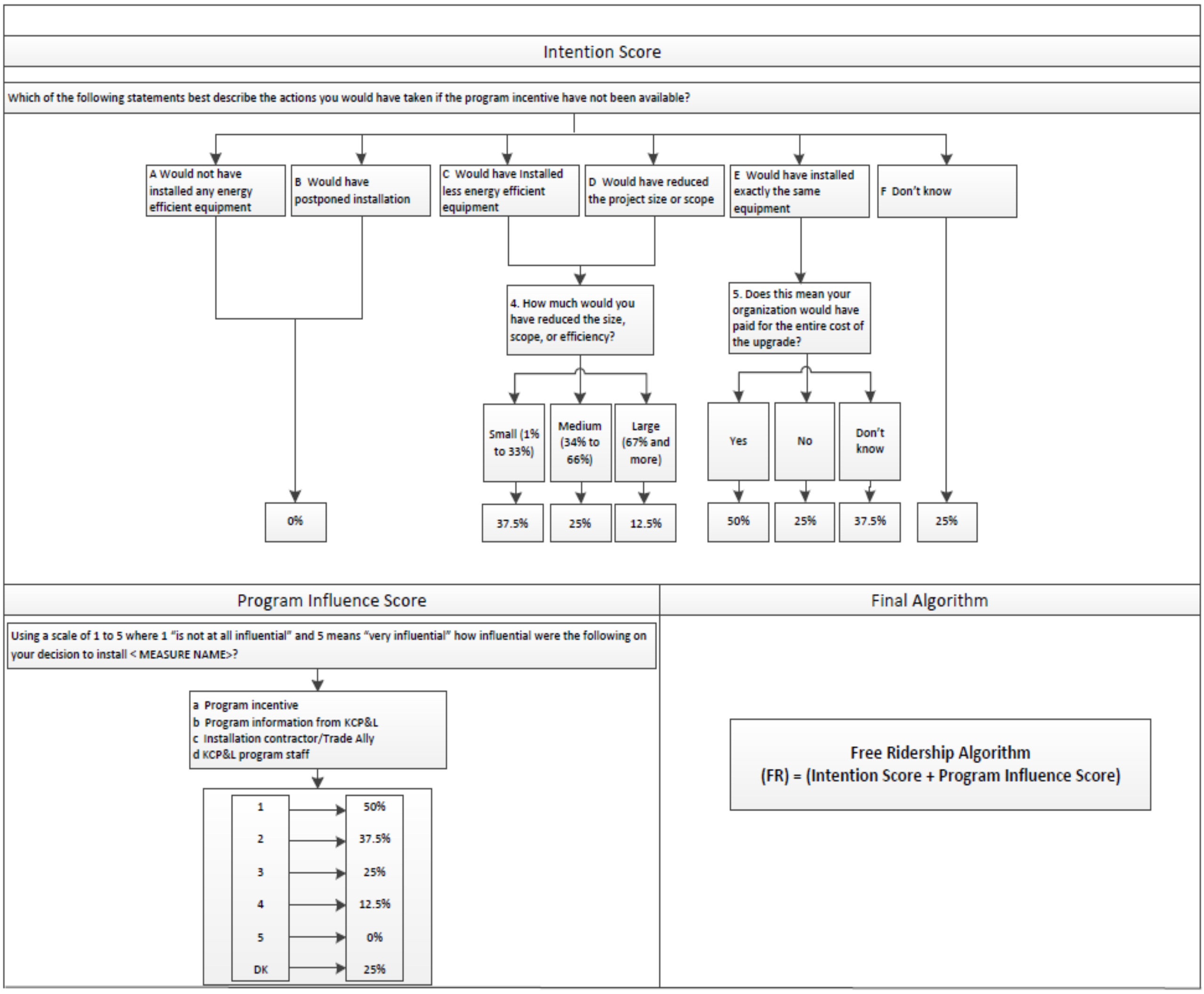
[9] Market sales analysis captures the total net effect of a program. Ideally, this method involves obtaining comprehensive pre- and post-market sales data in both the area of interest and in an appropriate comparison area and examining the change in the programme area compared with the change in the non-programme area (Tetra Tech et al. 2011).

[10] The New York Department of Public Service (2013a) presents guidelines for calculating the relative precision of programme net savings estimates for different types of estimates, including the NTG ratio based on the self-report method and for spillover savings.

Design of the Survey Based Attribution Analysis

This application used survey results to develop a behavioral model that is meant to capture the likelihood that participants would have installed some or all of the energy efficiency measures incented by the program, even if the programme had not existed. The approach was based on methodology based on Peters and Bliss (2013). The free ridership analysis included the following two elements: 1) intention to carry out the energy efficient project without programme funds, and 2) influence of the programme in the decision to carry out the energy efficient project. Figure 1 below illustrates the interaction between intent and influence of the program. Will the models developed using this framework be perfect? No, but it can provide an informed approximation that is not the same as a groundless guess (GHG Institute, 2012 p.26). Attribution and baseline models produce imperfect predictions of what actual behavior would have been in the absence of the programme, but these models and estimators can be reviewed and improved.

Figure 1. C&I Custom Programme Free Ridership Algorithm



Source: Navigant analysis drawn from Peters and Bliss (2014).

The methodology uses several elements of best practice design as set out in Violette and Rathbun (2014) to address potential biases and uncertainties in using surveys. The attribution rate is estimated using a triangulation approach. Information is collected from three sources for our analysis:

1. Programme participant fast feedback surveys (conducted soon after the customer's decision to participate),

2. Programme participant end-of-year telephone surveys, and
3. Programme Participant trade ally telephone interviews at end of year.

The FR and SO estimates are based on the most relevant information for that purpose:

1. Trade ally estimates of free ridership are used as a cap on the programme participant free ridership estimates.
2. Participating end-use customers are in the best position to articulate the likelihood that they would have been able to afford the high efficiency equipment without rebates.
3. Trade allies are best suited to comment on the influences of the programme beyond the rebate: for instance, the programme's influence on their technical knowledge, stocking patterns, and typical product specifications and recommendations.
4. Participant spillover (PSO) values are calculated on the basis of the end-of-year participant surveys.
5. Non-participant spillover value (NPSO) is calculated on the basis of the trade ally interviews.

These analyses were implemented for a commercial rebate programme run by an investor-owned utility in the Midwestern United States. The approach was peer reviewed by the expert EM&V auditor retained by the jurisdictional commission, and the work was also reviewed by stakeholders who represented customers, environmental interests, and jurisdictional regulatory authorities.

Table 1 and Table 2 below summarize Navigant's NTG research for these programs through end of year 2014,[11] with resulting free ridership and SO estimates represented by each survey.

Table 1 -- Preliminary NTG Component Results – C&I Custom Rebate Program

Survey Type	Respondents	Number	Fraction of Savings	Period	Free-Ridership	Participant Spillover	Non-Participant Spillover
Fast Feedback	Participants	114	23%	2014	0.14	--	--
End-of-Year	Participants	52	7%	2014	--	0.04	--
Trade Ally	Participating Trade Allies	20	56%	2014	0.19	0.005	0.11

Source: Navigant analysis

Table 2 -- Recommended Preliminary NTG Components and NTG Ratio

Program	Free-Ridership	Participant Spillover	Non-Participant Spillover	NTGR
C&I Rebate	0.14	0.04	0.11	1.01*

*Note: (NTG = 1 – FR + PSO + NPSO) -- *This value is estimated with 90% confidence and 9% precision based on protocols that assess the values of the survey responses if the entire population had completed the survey rather than just a sample.*

Participant NTG Component Analysis

[11] NTG component estimates have been estimated for 2015 and are under utility, stakeholder and regulatory review. Updates to these estimate should be available in time for presentation at the June IEPPEC conference.

This section presents Navigant’s analysis of NTG components derived from participant fast feedback and end-of-year surveys. Every C&I programme participant was contacted to participate in the fast feedback survey soon after their projects were approved in order to collect information on free ridership as close to the time of decision-making as possible. Navigant contacted 565 customers who received project approvals via phone or email. A total of 114 participants completed the fast feedback survey.[12] Each customer received the fast feedback survey approximately one month after project completion.

The telephone end-of-year survey was performed during the third quarter of 2014 and included respondents who completed projects during PY2014. The end-of-year survey collected information on both free ridership and spillover. The team interviewed a total of 52 customers.

Free Ridership. Navigant quantified free ridership from the participant surveys (both fast feedback and end-of-year) using the methodology presented in Figure 1. Navigant conducted a sensitivity analysis around the scoring approach used to determine the intention and influence scores used to develop the FR estimates. This addresses uncertainty in how self-reported questions on free ridership should be scaled, as described in the Uniform Methods Project chapter on net savings.[13] The results indicate that varying the scale has a small impact on the free ridership results from the fast feedback and end-of-year responses. However, more in-depth sensitivity analyses were proposed for the following year.

Spillover. Navigant quantified spillover from the end-of-year participant surveys. Navigant determined that 7 of the 52 respondents who participated in 2014 reported purchasing and installing additional energy-efficient equipment that was not rebated by the utility. Five of those seven respondents indicated the programme had an influence in the decision to purchase the additional equipment.

Trade Ally NTG Analysis

This section presents the analysis of NTG based on the results of telephone interviews with 20 participating C&I trade allies. Navigant designed a stratified sample for the trade ally interviews. The sample design divides the population of participating trade allies into three strata based on their 2014 programme savings. The stratified sample ensures good coverage of the very large trade allies, which represent a large share of programme savings (important for free ridership estimation) as well as medium and small trade allies (important for spillover estimation and process evaluation). Table 3 summarizes the interview sample design.

[12]This fast-feedback approach follows the approach used in Castor (2012).

Table3. Trade Ally Interview Sample Design

Strata	Number of Trade Allies in Population	% of Programme Savings in Strata	Targeted Number of Interviews	Actual Number of Interviews
Large Trade Allies (1,000,000+ kWh)	4	41%	4	3
Medium Trade Allies (200,000 to <1,000,000 kWh)	14	34%	6	7
Small Trade Allies (<200,000 kWh)	68	25%	10	10
Total	86	100%	20	20

Source: Navigant Analysis

The interviewed trade allies first answered a series of questions regarding the program's influence on aspects of the trade ally's business. These questions served two principal purposes:

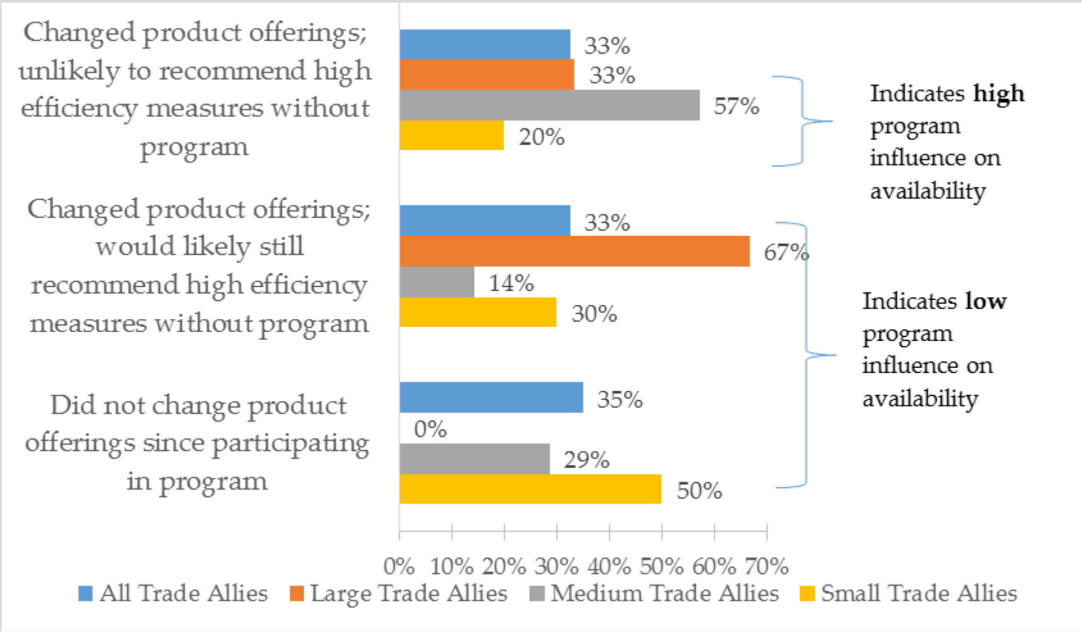
1. To remind trade allies of the various ways the programme may have influenced their business (beyond providing financial incentives to their customers).
2. To provide context to and a consistency check for the trade ally's direct estimate of free ridership and spillover.

The trade allies were then asked direct questions regarding the quantification of free ridership and spillover. The following sections detail these questions and Navigant's analysis of the results.

Programme Influence on Trade Allies (PITA). The interview included questions to assess the programs' influence on trade allies (i.e., PITA) since their participation with the program. These questions focused on programme influences which may not be noticed by end-use customers, including changes in the trade ally's typical stocking practices, product specifications, volume of high efficiency sales, and share of customers who choose high efficiency options. The responses to these questions provide context for the overall NTG assessment and serve as a checks to ensure that each trade ally's measure-specific free ridership and spillover responses are internally consistent with their responses (e.g., importance of program activities and estimates of free-ridership). Results of PITA questions are divided into two categories: **Influence on Availability** and **Influence on Project Volume**, both described below.

Influence on Availability. This metric represents the program's influence on the availability of high efficiency measure options. About two-thirds (65 percent) of the interviewed trade allies indicated that they had changed the efficiency levels of the measures that they offered to customers since participating in the program. For example, many of those trade allies indicated that they began offering LEDs or increased their offering of LEDs. Those trade allies who had changed their high efficiency options since participating were evenly split (50 percent/50 percent) on whether they would be likely or unlikely to recommend the same high efficiency options in the absence of the program. Altogether, these findings indicate that the programme had an influence on the availability of high efficiency options for approximately one-third of the interviewed trade allies, as shown in Figure 2.

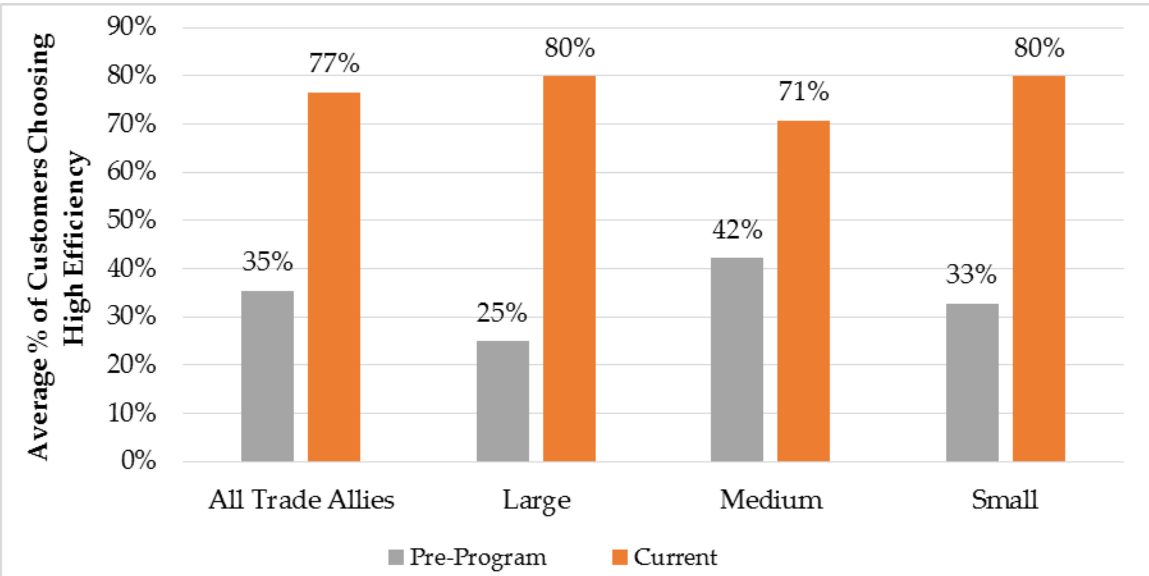
Figure 2. Programme Influence on Availability of High Efficiency Options (Trade Ally Survey)



Source: Navigant Analysis

Influence on Project Volume. This metric represents the program's influence on the number of high efficiency projects implemented by trade allies. All trade allies reported an increase in the overall number of high efficiency projects. Over three-quarters (80 percent) reported that their overall project volume (including high efficiency and standard efficiency projects) had increased. Nearly all (90 percent) reported that a higher percentage of customers choose high efficiency options now, relative to the pre-programme period. As shown in Figure 3, the average percentage of trade allies' customers who chose high efficiency options rose from 35 percent in the pre-programme time period to 77 percent currently. This increase was strongest for the largest trade allies interviewed.

Figure 3. Percentage of Customers Choosing High Efficiency Options (pre- and post-)



Source: Navigant Analysis

All of the trade allies report that the programme had a strong influence (4 or 5 on a 5-point scale) on the number of high efficiency projects they completed. This suggests that a good portion of the increase in the percentage of customers choosing high efficiency may be due to the programme and not solely due to other market factors. As stated before, the PITA results offer a valuable context in which to analyze free ridership and spillover (as well as offer a consistency check on individual trade allies' free ridership and spillover responses). Together, the contextual PITA results related to availability and project volume lead the Navigant team to expect a relatively high net-to-gross ratio (NTGR) for the programs. The individual components of the NTGR (free ridership and spillover) are discussed in the next two subsections.

Trade Allies' Direct Estimate of Free Ridership. After completing the series of questions focused on specific programme influences, trade allies answered measure-specific free ridership questions for their top two highest savings measures. Trade allies were asked about specific measures (e.g., replaced T12 with T8 lighting fixtures). However, for simplicity, the measures are grouped into three broad categories for the purposes of this analysis:

- LED lighting,
- other lighting measures (including controls), and
- non-lighting measures

Trade allies provided a direct estimate of free ridership as a response to the following question: "The programme database indicated that you had [# OF PROJECTS] projects in 2014 where [MEASURE NAME] was included and rebated by the programs. In your opinion, what percent of customers in these projects would have installed the same number of [MEASURE NAME] if the programme had not been offered?"

As a savings-weighted average, the trade ally estimate of programme free ridership is 19 percent. The interviewed trade allies represent over half (56 percent) of the total 2014 savings for the programs, indicating a robust sample design. The sample covers LED lighting well, with the 13 interviewed LED trade allies representing 73 percent of the programs' savings from LED lighting projects.

The trade ally free ridership estimates were found to be consistent based on an algorithmic check conducted by Navigant. This check ensured that respondents who estimated high free ridership responded to other questions in ways that indicate low programme influence and vice versa (i.e., low free ridership and high programme influence). However, the evaluation found very few inconsistent responses, and most of those related to the trade ally estimating high free ridership for one measure and low free ridership for another, indicating that the programme influence had more impact on some measures than others. Ultimately, no trade ally free ridership estimates required adjustment on the basis of the consistency check.

The final estimate of free ridership for the programs, as based on trade ally interview results, is 19 percent as presented in Table 4.

Table 4. Summary of Trade Ally Free Ridership Findings

Measure Category	% of Programme Savings in Measure Category	Number of Trade Allies Interviewed	% of Programme Savings Represented in Interview Sample[a]	Weighted Average Direct Free Ridership Estimate[b]
LED Lighting	55%	13	73%	0.20
Other Lighting/Lighting Controls	21%	11	42%	0.32
Non-Lighting Measures	24%	3	31%	0.05
Programme Level	100%	20[c]	56%	0.19[d]

a. *This column represents the interviewed trade allies' programme savings associated with the measure category relative to the total programme savings for that measure category; the programme level row represents the trade allies' savings as a percentage of all programme savings.*

b. *Each respondent's direct free ridership estimate was weighted by their share of the savings in the interview sample for that measure category, so that more active trade allies are weighted more heavily than smaller trade allies in the results.*

c. *Note that some respondents answered free ridership questions about measures in multiple measure categories; thus, the sum of the column exceeds the total.*

d. *The program-level free ridership is calculated by weighting each measure category's free ridership estimate by its share of the program's total savings (2nd column of the table).*

Source: Navigant analysis

Trade Allies' Estimate of Participant and Non-Participant Spillover. Interviewed trade allies answered a series of questions to establish the possible existence of spillover for their top three highest saving measures, as well as an open-ended question about other high efficiency measures that may result in spillover. The analysis looked for participant spillover (meaning non-incented high efficiency measures installed by customers who received programme rebates for other measures) as well as non-participant spillover (meaning efficient measures installed by customers who did not receive any programme rebates at all).

Nearly two-thirds (60 percent) of trade allies reported that they installed projects with high efficiency measures that did not receive rebates from the utility's program. For those trade allies who installed high efficiency measures out-side of the program, 58 percent (or 35 percent of the total) stated that the programme had influenced the non-incented projects, i.e., the savings from those non-incented measures can be at least partially attributed as spillover. Trade allies reported quantifiable spillover in four measure categories:

- LED lighting (replacing existing fixtures)
- T5 lighting (replacing High Pressure Sodium fixtures)
- Occupancy sensors
- T8 lighting (replacing T12 fixtures)

For each of these measures, Navigant used an **attribution factor** to calculate the **number of spillover projects.** This **attribution factor** is based on the trade ally's responses to programme influence questions. These trade allies indicated that they completed the vast majority of these spillover projects for non-participating customers (customers who had not received any programme rebates for any measures). Using the number of participant and non-participant projects, participant spillover is estimated as 0.5 percent and non-participant spillover is estimated at 11 percent of programme savings.

This trade ally spillover estimate is conservative for several reasons. First, the spillover estimate only accounts for projects implemented by six trade allies (i.e., those who noted the programme as an influential factor in non-programme projects) and it is does not extrapolate this result to the entire population of trade allies because of the high variability of the spillover results. This analysis assumed that the non-interviewed trade allies' (participant and non-participant) spillover is zero, which is unlikely.[14]

While the survey did not directly ask trade allies why spillover equipment was not part of the program, some trade allies indicated that some customers think the application process is too involved and that the rebate is not worth the extra effort.

Reconciliation of Participant and Trade Ally NTG Results. Navigant recommends using the trade ally free ridership estimate as a cap on the participant free ridership estimates. Participating end-use customers are in the best position to articulate the likelihood that they would have been able to afford the high efficiency equipment without rebates. Trade allies are best suited to comment on the influences of the programme beyond the rebate: for instance, the program's influence on their technical knowledge, stocking patterns, and typical product specifications and recommendations. The participants are often unaware of how these non-rebate programme influences may have shaped their experiences with the trade ally, and thus they may be prone to overestimating free ridership in self-report surveys. However, since the participant free ridership is lower than the trade ally free ridership estimate, there is not a compelling reason to believe that the participants have misjudged the program's influences on their purchases. Therefore, with the trade ally free ridership estimate as a cap, Navigant recommends using the participant free ridership estimate to calculate the final NTG ratio.

With respect to participant spillover, the value calculated on the basis of the end-of-year participant surveys and the non-participant spillover value calculated on the basis of the trade ally interviews are used. Thus, Navigant recommends excluding the trade ally participant spillover estimate from the analysis to avoid double counting participant spillover. Navigant recommends simply excluding the trade ally estimate of participant spillover, rather than averaging or blending it with the participant survey results, because trade allies inherently have an incomplete view of participant spillover -- particularly unlike spillover that is outside the measures offered by the program and if participants are purchasing additional, non-incented high efficiency measures from other trade allies, retailers, or distributors.

Conclusions

The survey-based attribution application presented utilized a number of best practices as developed in the literature. Like all attribution analyses, it is a based on a model that is designed to capture important influences of the programme on EE adoption. The problem of attribution is predicated on developing information on what would have occurred absent the program. This counter-factual scenario can never be directly observed. As a result, the goal is to produce an informed estimate that can help validate that the savings are in fact additional, i.e., are actually attributable to the EE programme. Whether this meets the burden of proof required by stakeholders and provides the information needed to make good decisions regarding EE programme designs and investments will be determined in a policy context (Violette and Vantzis, 2014). Many important investment decisions are made with less reliable information. Estimating and determining attribution within an EE policy framework consistent with that jurisdiction's goals and views regarding EE objectives can be challenging. The approach and use of the estimates can be context dependent. That is, one jurisdiction with one set of policies may use savings estimates differently than another jurisdiction, yet both can be appropriate and consistent given their respective overall sets of EE policies and objectives.

[14] Extrapolating the non-participant spillover estimate to the population of participating trade allies would result in an estimate of NPSO of 17% of programme savings, which would be the high estimate. Navigant will explore other methods for extrapolating surveyed trade ally responses to the population to determine an appropriate NPSO at the end of PY2015.

The costs of evaluation and different approaches to assessing attribution has been an important element of the discussion. Concerns about the cost and practicality of survey-based methods were expressed in both the Workshop held in Brussels (Austrian Energy Agency, 2015), and also in Ecofys (2012) with the implication that survey-based methods may be more expensive relative to the reliability of the methods.[15] Survey-based methods are more often used in North America and several robust studies have been conducted in Denmark (Bundgaard et al., 2013 and Togeby et al., 2009). However, experience with survey-based evaluation methods in North America has shown that they can be conducted at reasonable costs and quite possibly at lower costs that some BAU methods.

The BAU (or common practice baseline) approach has not been advocated as applicable to all programs, even within a single jurisdiction in North America (Hall et al., 2013). An evaluator can select from among the many other methods for estimating net savings, each with its own sources of error, and decide which is most likely to produce the highest quality information given the cost of the research.

Several jurisdictions in North America are looking toward the use of BAU baselines in their EE evaluation guidelines. As with all methods, there are pros and cons. A potential strength of the BAU approach is its use in upstream and market transformation EE programs. It can be applied market-wide and, unlike randomized trials and quasi-experimental designs, it does not require participants to be identified if appropriate sales data are available (and that is a big IF). However, this method is more susceptible to self-selection issues (that is, the average consumer may not be the type of consumer who participates in the program). It is not clear how this can be addressed, other than by conducting surveys to determine specific characteristics of purchasers of efficient equipment relative to the common practice baseline. However, this survey effort would negate the unique aspects claimed for the common practice baseline approach; i.e., specific consumers who have and have not purchased the high efficiency equipment would need to be identified. This makes this approach more similar to the survey-method attribution approach discussed in this paper.

SEE Action (2012b, p. 7) indicates that appropriate BAU and common practice baselines can be estimated through surveys of participants and nonparticipants as well as analysis of market data. The process of developing a working definition of BAU approaches and common practice baselines may pose some challenges. In North America, there is not widespread experience in developing common practice baselines allowing for a determination of best practices. In addition, some BAU baselines will be more difficult to develop than others. Custom rebate programs in the C&I sector pose challenges due to the variety of equipment that may be installed and the fact that there may be a smaller set of customers in the market installing the same types of equipment. Some of these issues may be more easily addressed in residential programmes and measures offered; however, programmes target at the residential sector are also more amenable to other methods. In particular, random control trials have become more common for certain residential programmes with large number of programmes, and where randomization is not possible, advanced matching methods are being used to develop matched comparison groups. The use of these new methods are providing increased confidence in estimates of additional savings from statistical evaluation methods for residential customers with a large number of participants.

In general, there have been questions about the cost and complexity of BAU methods versus traditional methods. All methods pose challenges, and costs will vary across programme type and sector. For example, it can be difficult to gather data for determining the BAU baseline if there are a large number of measures across different types of customers. There is also a need to update BPU baselines as needed. Attempts to gather actual EE measure market sales data have proved difficult, and extrapolation from a small base of sales data to the jurisdiction or service territory can be uncertain. Trade ally panels can be used to develop expert-based judgmental baselines. However, the scale of the effort is important to assess. There may be a large number of measures and practices that need to have BAUs developed and

[15] "The trade-off between the predictability, certainty of additionality and the administrative cost of the scheme is a political decision. An effective and robust approach to M&V and additionality of electricity savings is always a trade-off between the costs of the approach and the certainty of the savings achieved. Electricity" Ecofys (2012, p.24).

updated with reasonable frequency. It is not clear that a general statement can be made about whether BPU methods are any more or less resource intensive than other methods and there are threats to validity with all methods. A BAU method may draw heavily on engineering method and be more data driven and, therefore, may appeal to those evaluators whose background is in these areas. However, are these methods more accurate in terms of producing estimates of additionality closer to the true values?

Finally, getting out into the field and engaging program participants and trade allies is an important part of EE programme evaluation. It is important to talk with participants and trade allies to get at least rough estimates of FR and SO periodically to confirm whether they are likely to be large and need more attention either from programme design, or from an attribution perspective.

References

Austrian Energy Agency (2015). "Energy Saving Policies and Energy Efficiency Obligation Scheme D3.2: Report Workshop on Article 7 of the Energy Efficiency Directive." Project Coordinator: Joint Implementation Network – JIN. Task 3.2 - Leader Organization: Austrian Energy Agency March 2015

Baumgartner, R. (2013). "Survey Design and Implementation Cross-Cutting Protocols for Estimating Gross Savings." Chapter 12 in *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*. www.nrel.gov/docs/fy13osti/53827.pdf.

Bundgaard S.S., M. Togeby, K. Duhr-Mikkelsen, T. Sommer, V. Hansen-Kjaerbye, and A. Larsen. 2013. "Spending to Save: Evaluation of the Energy Efficiency Obligation in Denmark". *In Proceedings of the ECEEE 2013 Summer Study on Energy Efficiency*. Toulon, France. European Council for an Energy Efficient Economy.
http://proceedings.eceee.org/visabstrakt.php?event=3&doc=7-131-13.

Castor, S. (2012). *Fast Feedback Results.* 2011 Final Report prepared for Energy Trust of Oregon. http://energytrust.org/library/reports/Fast_Feedback_-_20110.pdf

Ecofys (2012). Measurement, Verification and Additionality of Electricity Demand Reductions Final report – recast. Prepared for: DECC (Department of Energy and Climate Change). By: Isabelle de Lovinfosse, Luis Janeiro, Kornelis Blok, Julia Larkin-Date: 6 November 2012, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/66216/Electricity_demand_reduction-_measurement__verification_and_additionality.pdf

GHG Institute (2012) "What is Additionality? Part 2: A framework for more precise definitions and standardized approaches. By Michael Gillenwater, Discussion Paper No. 002 January. http://ghginstitute.org/wp-content/uploads/2015/04/AdditionalityPaper_Part-2ver3FINAL.pdf

Hall, N; Ladd, D.; Khawaja, M.S. (2013). "Setting Net Energy Impact Baselines: Building Reliable Evaluation Approaches." Paper presented at the 2013 International Energy Program Evaluation Conference, Chicago, IL.

Keating, K. (2009). "Free ridership Borscht: Don't Salt the Soup." Paper presented at the 2009 International Energy Program Evaluation Conference.
Navigant (2013). *Custom Free Ridership and Participant Spillover Jurisdictional Review*. Prepared for the Sub-Committee of the Ontario Technical Evaluation Committee, May.

www.ontarioenergyboard.ca/documents/TEC/Evaluation%20Studies%20and%20Other%20Reports/Ontario%20NTG%20Jurisdictional%20Review%20-%20Final%20Report.pdf.

NEEP (2010). *"Net Savings Scoping Paper."* Revised Draft prepared for Northeast Energy Efficiency Partnerships: Evaluation, Measurement, and Verification Forum. By: NMR Group, Inc. and Research Into Action, Inc. www.neep.org/Assets/uploads/files/emv/emv-products/FINAL_Net_Savings_Scoping_Paper_11-13-10.pdf

NEEP (2016). "Gross Savings and Net Savings: Principles and Guidance, prepared for the NEEP" Regional EM&V Forum. By Daniel Violette, Pam Rathbun, and Teresa Lutz. Available at NEEP website: www.neep.org.

New York Department of Public Service (2013a). *Guidelines for Estimating Net-To-Gross Ratios Using the Self-Report Approach*. Appendix H.https://www3.dps.ny.gov/W/PSCWeb.nsf/96f0fec0b45a3c6485257688006a701a/766a83dce56eca35852576da006d79a7/$FILE/NY_Eval_Guidance_Aug_2013.pdf

New York Department of Public Service (2013b). *Guidelines for Calculating the Relative Precision of Program Net Savings Estimates*. Appendix I.https://www3.dps.ny.gov/W/PSCWeb.nsf/96f0fec0b45a3c6485257688006a701a/766a83dce56eca35852576da006d79a7/$FILE/NY_Eval_Guidance_Aug_2013.pdf

Peters, J. and R. Bliss (2014). "Common Approach for Measuring Free-riders for Downstream Programs", Appendix F. For the Statewide Evaluation (SWE) for the Pennsylvania Electric Distribution Companies (EDCs) and the Pennsylvania Technical Utility Staff (TUS), December 23, 2014. http://www.puc.pa.gov/Electric/pdf/Act129/SWE_PhaseII-Evaluation_Framework060114.pdf

Ridge, R.; Willems, P.; Fagan, J.; Randazzo, K. (2009). "The Origins of the Misunderstood and Occasionally Maligned Self-Report Approach to Estimating Net-to-Gross Ratio." Paper presented at the 2009 Energy Program Evaluation Conference, Portland, OR.

SEE Action (2012). *Energy Efficiency Program Impact Evaluation Guide*. Prepared by Steven R. Schiller, Schiller Consulting, Inc. See: www1.eere.energy.gov/seeaction/pdfs/emv_ee_program_impact_guide.pdf.

Slote S., M. Sherman, and D. Crossley. 2014. *Energy Efficiency Evaluation, Measurement, and Verification*. Regulatory Assistance Project (RAP) Global Power Best Practice Series. http://www.raponline.org/topic/global-power-best-practice-series.

Togeby, M., et al. 2012. "A Danish Case Study: portfolio evaluation and its impact on energy efficiency policy." *Energy Efficiency* 5: 37–49. http://link.springer.com/article/10.1007%2Fs12053-011-9117-7#page-1.http://link.springer.com/article/10.1007%2Fs12053-011-9117-7#/page-1

Tetra Tech (2011). *Cross-Cutting Net to Gross Methodology Study for Residential Programs – Suggested Approaches*. Final report prepared for the Massachusetts Program Administrators. http://ma-eeac.org/wordpress/wp-content/uploads/Cross-Cutting-Net-to-Gross-Methodology-Study-for-Residential-Programs-Suggested-Approaches-Final-Report.pdf

Violette, D. and B. Rogers (2012). *A Sampling Methodology for Custom C&I Programs*. Prepared by Navigant Consulting, Inc., prepared for the Ontario Natural Gas Technical Evaluation

Committee, Ontario Energy Board, November. See: report available at Ontario Energy Board website:
http://www.ontarioenergyboard.ca/documents/TEC/Evaluation%20Studies%20and%20Other%20Reports/TEC%20SC%20-%20Sampling%20Method%20-%20Final%20Report%2020121112.pdf

Violette, D. M. and P. Rathbun. "The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures, Chapter 23: Estimating Net Savings: Common Practices," DOE UMP 2014. National Renewable Energy Laboratory (NRLEL), September 2014. http://energy.gov/sites/prod/files/2015/02/f19/UMPChapter23-estimating-net-savings_0.pdf.

Wuppertal Institute (2009) "Measuring and reporting energy savings for the Energy Services Directive – how it can be done." Results and recommendations from the EMEEES project. On behalf of the EMEEES Consortium.http://www.evaluate-energy-savings.eu/emeees/en/publications/reports/EMEEES_Final_Report.pdf

.