Meaningful Impact:

Challenges and Opportunities in Industrial Energy Efficiency Program Evaluation

Anna Chittum

September 2012

Report Number IE122

529 14th Street NW, Suite 600, Washington, DC 20045

Phone: (202) 507-4000 • Twitter: @ACEEEDC

Facebook.com/myACEEE • www.aceee.org

Contents

Executive Summary iii

Acknowledgments iv

Glossary and Suggested Definitions v

Introduction 1

Methods 3

Why We Evaluate 4

- The Nature of Ratepayer-Funded Energy Efficiency 4
- Ensuring Prudent Use of Funds 5
- Ensuring Reliable Energy Savings 6

Program Evaluation 7

- What Evaluation Should Reveal 7
- Components of Evaluation 8
- Differences in Evaluation Administration 10
- How Measurement and Verification Informs Evaluation 12
- How Evaluation Findings Are Used 13

The Challenges of Industrial Program Evaluation 14

- The Nature of Industrial Facilities 15
- Evaluation's Players 16

Six Major Issues in Industrial Program Evaluation 17

- Developing a Baseline 18
- Timing of evaluation activities 20
- Net Savings and Net-to-Gross Ratios 22
- Free Riders 28

Spillover 32

Non-Energy Benefits 33

The Future of Industrial Program Evaluation 36

Attribution and Climate Change 36

Energy Management Activities 37

Self-Direct and Opt-Out Programs 38

Discussion 39

The Role of Regulators 39

Conclusions and Next Steps 40

References 43

Appendix I: Telephone Interview Framework 47

Executive Summary

Energy efficiency programs run by utilities and public benefit fund (PBF) organizations must meet certain energy savings goals, cost-effectiveness tests, and other requirements to ensure that the programs are spending public dollars effectively, fairly, and prudently. Impact evaluations are the assessments that typically ascertain the degree to which programs are impacting the energy savings of their program participants. The manner in which these programs are evaluated for their impacts and their success in meeting relevant goals differs from state to state.

Evaluation in any sector is difficult. Evaluation aims to measure something that does not exist—the energy that is *not* being used—and so evaluators must rely on assumptions and estimates to make their determinations. Programs that acquire energy efficiency within the industrial sector are evaluated to different degrees of stakeholder satisfaction. The unique attributes of the industrial sector, and the way the industrial sector makes decisions about energy efficiency investments, present unique challenges to the effective evaluation of industrial energy efficiency programs.

Based on interviews and surveys with industrial energy efficiency program managers, evaluators, and regulators, this report discusses how industrial energy efficiency program evaluation is conducted and the types of data and metrics derived by evaluators. It explains the use of these various metrics and the manner in which specific metrics are developed.

Evaluation of industrial energy efficiency programs is a necessary activity to ensure public funds are used in a prudent and responsible manner. However, evaluation activities may not always reflect how industrial energy efficiency programs acquire energy efficiency and how customers make decisions about their energy efficiency investments. Additionally, while evaluation activities can yield useful information that can help improve existing industrial energy efficiency programs, there are changes that could be made to improve evaluation activities and make the findings more useful to efficiency program designers and implementers.

This report discusses six issues in-depth that were of particular interest to interviewees and survey respondents. They are:

- The development of a facility's baseline
- The timing of evaluation activities
- The measurement of net savings and the use of net-to-gross ratios
- The measuring of free riders and their associated savings
- The measurement of spillover effect
- The measurement of non-energy benefits

This report finds that there are substantial concerns about and varied perspectives on all of the above issues. Evaluators, regulators, program administrators, and industrial customers all believe that many of the above components of evaluation are insufficiently or inaccurately being conducted. This report explains the concerns about each issue and suggests best practices and recommended directions for improvement where available and applicable. One important finding is that best practices are difficult to find in industrial energy efficiency program evaluation because few evaluation approaches are considered best practices by all stakeholders. An additional finding is that the emphasis on high levels of precision within program evaluation activities may not always be justified or useful when other types of assessments would suffice or where such precision is not achievable.

While substantial room for improvement exists in most current evaluation approaches, there are indeed some instances in which program evaluation is evolving and improving to better meet stakeholder needs. However, industrial energy efficiency programs are positioned to grow and change in the near future by encouraging stronger energy management practices and working to improve industrial facilities' energy performance. This report also finds that current approaches to program evaluation may not be well-suited to the industrial energy efficiency programs of tomorrow. .

Future industrial energy efficiency programs will more tightly integrate resource acquisition and market transformation efforts seamlessly into customer engagement. Such programs may be too nuanced for some of the evaluation methodologies in place today. So too will future emissions-reduction programs overlap with the activities of energy efficiency programs. Discerning among influences will likely become more difficult. Evaluation will continue to be a critical component of industrial energy efficiency programming, but it will need to adapt to these changes to yield findings that are useful and meaningful to all parties.

Acknowledgments

The author would like to thank the various funders for their generous support of this research, including the New York State Research and Development Authority, Pacific Gas and Electric Company, and Southern California Edison; every program administrator, evaluator, regulator, and policymaker interviewed and surveyed for their candid and insightful comments provided throughout the research process; Neal Elliott, Ethan Rogers, Marty Kushler, and Steven Nadel of ACEEE for their feedback and guidance throughout the report; and Renee Nida, Patrick Kiker, Eric Schwass, and Glee Murray of ACEEE for their help in producing and disseminating the final product.

Glossary and Suggested Definitions

a. Adjusted gross savings
    i. Gross savings that have been adjusted to reflect the applicable realization rate.
b. Baseline
    i. Energy consumption that would have occurred without implementation of the energy efficiency measure, project, or service. Often referred to as business-as-usual (adapted from NMR 2010).
    ii. Least efficient, non-regressive, code or regulations-compliant option specific to a particular facility and application that the customer technically, functionally, and economically could have alternatively considered to deliver the post-retrofit level of production or service (Maxwell et al. 2011).
c. Deemed savings
    i. Savings estimates for a specific measure based on established and universally accepted fixed measure-specific savings estimates.
d. Evaluator
    i. An individual or organization tasked with the evaluation of an energy efficiency program. Most often not a member of the organization administering the energy efficiency program, although internal evaluators within program-administering entities such as utilities and PBF organizations do exist.
e. Free-rider
    i. A program participant who would have implemented the measures or practices in question absent the program (NAPEE 2007 and NMR 2010).
f. Gross energy savings
    i. The changes in energy consumption and/or energy demand seen in customers that have been exposed to an energy efficiency program, regardless of why they participated. The "physical change in energy use after taking into account factors beyond the customer or sponsor's control (e.g., weather)." (NAPEE 2007 and NMR 2010).
g. Impact evaluation
    i. Evaluation activities that determine the actual savings and benefits of the given energy efficiency program.
h. Net energy savings
    i. The energy savings that would not have occurred absent the energy efficiency program (NMR 2010). Sometimes calculated as gross savings minus free rider savings plus spillover savings.
i. Net to gross ratio
    i. Net program savings divided by gross program savings. This ratio is then typically applied to gross program impacts to derive net program impacts (NMR 2010).

j. Non-energy benefits
    i. Benefits other than energy (kWh, kW, therms) that accrue to program participants. These can be specific resources, such as water savings, or slightly less tangible things, such as improved worker safety and comfort.
k. Program Administrator Cost Test (PACT)
    i. A cost test that compares a utility's or program's avoided cost benefits with energy efficiency program expenditures (ECW 2012).
l. Program manager or administrator
    i. An individual or organization responsible for the deployment of all or part of an energy efficiency program.
m. Public benefit fund (PBF) organization
    i. An entity tasked with the prudent use of collected public benefit or system benefit fees contributed by energy users. PBF organizations administer energy efficiency and renewable energy programming in several states in the U.S.
n. Realization ratio
    i. Evaluated and verified energy savings results divided by initial estimated energy savings (adapted from NMR 2010).
o. Spillover
    i. Spillover energy savings occur when a program, by its mere presence, indirectly influences measures or behaviors that save energy but the customers or facilities implementing those measures or behaviors are not actual program participants. One type of spillover, called "inside spillover," occurs when program participants implement changes elsewhere in their facilities as a result of being a program participant.
p. Societal Cost Test (SCT)
    i. A cost test that includes all of the costs and benefits address in the TRC, as well as external benefits to society such as avoided emissions.
q. Total Resource Cost test (TRC)
    i. A cost test that compares the energy avoided cost benefits to society with the participant's cost of installing the measure plus the cost of energy efficiency program administration (ECW 2012).

Introduction

The focus of this report is industrial energy efficiency program impact evaluation. Impact evaluations help determine the impact an energy efficiency program is having on the energy performance of its customers. These evaluations are then used by regulators, policymakers, and the programs themselves to make conclusions about program performance and decisions about future program development.

Though industrial energy efficiency programs and their evaluations look quite different from program to program, generalizations can still be made. Throughout this report, "program" will refer to an energy efficiency program focused on or including industrial sector customers. "Program administrators" and "managers" are used to describe individuals or organizations tasked with the actual operation and deployment of energy efficiency programs. "Evaluators" will be used loosely to describe individuals and organizations tasked with conducting a program's impact evaluation.

This report relies on primary and secondary data. The primary data is derived mostly from telephone interviews and an online survey administered to industrial energy efficiency program administrators. The secondary data largely comprises evaluation reports, public service commission documents, and academic articles from experts in the evaluation community. The goal of this data collection was to identify current trends and challenges in industrial program evaluation through the eyes of a variety of stakeholders. Impact evaluation is very important in determining the success and influence of a program, but many stakeholders appear to be unhappy with the manner in which such evaluations are currently conducted. This report attempts to summarize those frustrations and issues, and suggest possible ways of improving industrial energy efficiency program evaluation.

Evaluating the impact of energy efficiency programs is critical, because policymakers are increasingly relying on energy efficiency as a primary energy resource. Energy efficiency is the cheapest, most quickly deployable new energy resource available to states today, and industrial energy efficiency is very often the most cost-effective efficiency resource within a utility or public benefit fund (PBF) organization's[1] entire efficiency portfolio (Friedrich et al. 2009, Chittum and Nowak 2012).

Recent research discovered that approximately 84% of all funding for industrial energy efficiency programs in the United States is administered by utilities and PBF organizations, with the balance provided by government entities, universities, and nonprofit organizations (Chittum and Nowak 2012). These utility and PBF programs thus represent the most significant current efforts to acquire industrial energy efficiency resources in the U.S. today. It is imperative, then, that their performance is measured in a meaningful way.

[1] States frequently collect public (or "system") benefit charges from utility ratepayers and set aside those moneys to pay for energy efficiency or renewable energy resources that benefit the entire system. A "public benefit fund organization" is an entity tasked with the prudent deployment of those funds. Public benefit fund organizations take many forms, but the most salient characteristics are that some or all of the system benefits funds collected from ratepayers are given to the public benefit fund organization, and that the organization in turn must use those funds responsibly, in a cost-effective and equitable manner.

ACEEE has a long history of working with the program administrators of these programs and learning about current challenges and issues in program deployment. One common issue noted by program administrators was that the manner in which programs are evaluated for their impacts and their success in meeting relevant goals differs significantly from state to state. While some programs appeared to be neutral about the way in which their programs were evaluated, most indicated some degree of frustration with the evaluation process, the outcomes, the metrics, or some combination of the three.

Industrial energy efficiency investments are particularly challenging investments to accurately evaluate. Evaluation should yield information about what influenced energy efficiency investments, but many industrial efficiency investments were influenced by multiple parties, and calculating specific influence can be very hard or even impossible. Evaluation should yield some understanding of energy savings over a long period of time, but industrial energy efficiency measures and facilities may perform very differently from week to week and season to season. Evaluation should also determine the cost-effectiveness of various measures and programs, but some industrial facilities find the non-energy benefits of investing in a measure to exceed the actual energy benefits, even though the program may not fully measure non-energy benefits. In short, the industrial sector is complex and industrial programs are often difficult to evaluate.

The challenges are only growing. With new concerns about the relationships between energy efficiency programs and larger efficiency or climate-related goals, ACEEE determined that now was a good time to visit the issue of industrial energy efficiency program evaluation and understand current concerns held by stakeholders.

One of the great challenges in comparing and contrasting different approaches to program evaluation is that each state, utility, or PBF organization often has its own definition of major evaluation terms and concepts (Kushler et al. 2012). Evaluation activities can include a number of different data collection and analysis efforts. Depending upon how a program tracks customer energy use and savings, evaluation approaches can involve a variety of tools. The Glossary of this report lists definitions of many of these major terms and components of evaluation.

This report summarizes and explains the major components of industrial energy efficiency program impact evaluation activities, and addresses some of the critical challenges to better and more useful program evaluations. It discusses concerns held by a variety of stakeholders, such as the definition and calculation of free riders, non-energy benefits, and spillover effects. It explores some of the new challenges presented by changes in the way industrial energy efficiency programs acquire energy efficiency in the industrial sector. It also identifies some best practices in key areas of industrial energy efficiency program evaluation.

This report finds that, while evaluation has a long and solid reputation for providing unbiased analysis of program impacts, most program administrators feel that evaluation could better or more accurately analyze their program impacts. Concerns about particular metrics, such as free riders and net savings, and the approaches to deriving those metrics were voiced by a wide variety of stakeholders. Those concerns are presented throughout the report, along with suggestions of how they could be addressed.

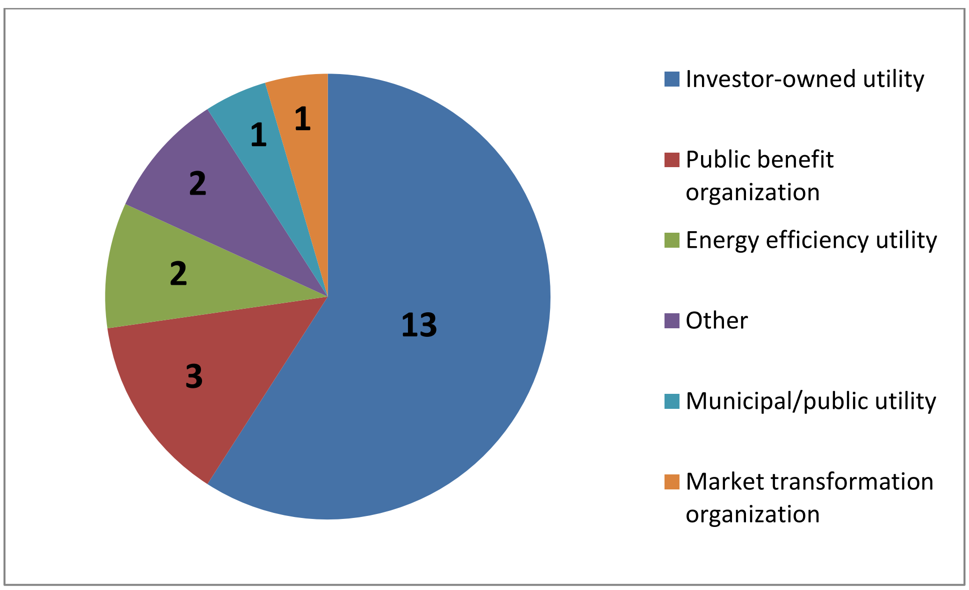
This report concludes by identifying some critical questions that must be asked of evaluation activities, especially as industrial energy efficiency resources become more important to certain states and utilities. Changes in how industrial energy efficiency is acquired will necessitate changes in program evaluation, and this report highlights some potential changes on the horizon for both industrial energy efficiency programs and the evaluators who analyze their impacts.

Methods

This report relies primarily on two types of primary data: informational interviews, conducted via telephone with about one dozen industrial energy efficiency program administrators, program evaluators, or program regulators; and an online survey ("the online survey") of 25 industrial energy efficiency program managers. Additional online surveys of five state utility regulatory commissioners and telephone interviews with several established experts in the field of program evaluation helped shape a national picture of industrial energy efficiency program evaluation.

The online surveys were conducted anonymously, and respondents were told their answers would not be attributed to their organizations. Results of the survey are reported in the aggregate throughout this report. The types of entities to which survey respondents belonged are shown in Figure 1.

Figure 1. Number of Survey Respondents per Type of Organization



Notes: Three additional respondents participated but did not indicate their affiliation.
Source: Online survey

More information about the telephone interviews can be found in Appendix I. Additionally, Kushler et al. 2012, an ACEEE report published earlier this year, offers detailed information about each state's approach to energy efficiency program evaluation, though not specific to *industrial* programs.

Why We Evaluate

There are several reasons why energy efficiency programs are evaluated. These include, as identified by interview subjects and published evaluation reports:

- To measure energy savings and other impacts of specific programs, especially for system planning purposes;
- To determine the performance of programs, especially the cost-effectiveness of them;
- To measure environmental impacts of programs;
- To attribute specific savings to certain programs;
- To identify areas in which programs can be improved; and
- To make some high-level comparisons between programs of different types, sizes, and scopes.

Energy efficiency programs are often accountable to a large number of people, including ratepayers and taxpayers. The nature of their funding necessitates significant oversight, usually conducted via evaluation.

THE NATURE OF RATEPAYER-FUNDED ENERGY EFFICIENCY

Energy efficiency programs run by utilities and PBF organizations are typically funded by the same people who benefit from increased energy efficiency: ratepayers (i.e., customers of the local utility). Ratepayers pay for energy efficiency programs through public benefit funds, energy efficiency riders or surcharges, or through additional fees embedded in their utility rates. These funds are then set aside specifically to fund energy efficiency programs. Utilities or PBF organizations are tasked with the responsible use of these funds and the administration of energy efficiency programs. Currently, 44 states deploy some type of ratepayer-funded mechanism to pay for energy efficiency programming (Chittum and Nowak 2012, Kushler et al. 2012).

Energy efficiency programs serve certain segments of a utility or PBF organization's customers, often offering specific programs for residential, commercial, and industrial customers. However, industrial customers are often served concurrently with commercial customers, and only 40% of respondents in the online survey indicated that they offer distinct industrial programs with dedicated staff to their industrial customers. A full 28% of respondents indicated that industrial customers are treated as commercial customers, with no distinction in program offerings or staff.

Experience shows that industrial customers feel better served by dedicated industrial programming and staff than programs designed to serve industrial and commercial facilities jointly (Chittum et al. 2009). While this report will not explore the details of specific program offerings, energy efficiency programs for industrial customers or industrial and commercial customers often include some or all of the following:

- Energy audits, assessments, or other technical assistance to identify potential energy efficiency projects;
- Financing or financing assistance for identified projects;
- Financial incentives or rebates for specific energy performance or technologies;
- Recognition of energy-efficient performance or energy savings;

- Energy management system deployment and support;
- Training and education;
- Incentives for new facility construction; and
- Standard offer programs, which offer a fixed financing benefit for a contracted amount of energy savings (Chittum et al. 2009).

Industrial customers often utilize several of these offerings at once, combining incentives and technical assistance to implement projects that improve overall energy efficiency while yielding other benefits such as improved safety, improved production efficiency, or reduced onsite emissions.

Ensuring Prudent Use of Funds

Energy efficiency programs administered by utilities or PBF organizations are in general designed to acquire cost-effective energy efficiency resources that offer long-term economic and environmental benefits to all system users.[2] Ratepayer dollars are watched keenly by regulators and lawmakers because concerns about extra "costs" on consumers' bills mean negotiating energy efficiency program funding is often a contentious process. Evaluation should, if programs are being run well, reveal the overall benefit of looking to energy efficiency as an energy resource.

Evaluation can help determine how beneficial certain efficiency programs are. Some energy efficiency investments are much more cost-effective than others, and evaluation can determine the cost-effectiveness of specific measures or programs. Some investments would happen anyway, regardless of the activities of a utility or PBF organization's energy efficiency programming, and this effect, called *free-ridership*, is often measured by evaluation. Additionally, some energy efficiency programs end up encouraging energy-efficient behavior in facilities and customers not directly involved in the energy efficiency program. This effect, called *spillover*, is also sometimes measured by evaluation. These effects will be examined in much greater detail later in the report.

Ratepayers, regulators, policymakers, and the utility or PBF programs themselves want to understand the degree to which their programs are actually saving energy, and whether other external forces might be causing some facilities to make energy efficiency investments. They also want to know whether certain program offerings, such as specific incentives, are truly still needed in the marketplace. If a certain technology has become standard practice, utilities and PBF programs may decide to no longer offer an incentive for that technology and to instead redirect those funds into a different offering.

Understanding all of these characteristics and impacts of an energy efficiency program can help define the program's value and determine whether it is contributing to a cost-effective portfolio of energy

[2] Other energy efficiency programs are designed to encourage market transformation, but these are typically evaluated differently and are not the focus of this report. *Market transformation* is the strategic process of intervening in a market to create lasting change in market behavior by removing identified barriers or exploiting opportunities to accelerate the adoption of all cost-effective energy efficiency as a matter of standard practice. For additional information, visit: http://www.aceee.org/topics/market-transformation.

efficiency resources. All states with ratepayer-funded energy efficiency use some sort of cost-based test or analysis to determine which projects should be supported by its programs (Kushler et al. 2012). Most states use the total resource cost (TRC) test, which compares the overall energy avoided cost benefits of a measure to the cost incurred by the program participant and the program itself, minus incentives (ECW 2012). States also frequently use societal cost tests (SCT) and program administrator/utility cost tests (PACT), which include additional societal externalities and costs to utilities, respectively (Kushler et al. 2012, ECW 2012). Evaluation activities show how well programs are satisfying these cost tests and whether they are conveying the desired benefits to all a system's users.

One notable exception to the above is opt-out and self-direct programs, which are offered to certain large industrial and commercial customers in about half of U.S. states. These programs are explained and explored in detail in Chittum 2011. Opt-out provisions allow qualifying customers to opt out of paying PBF funds and have no follow-up to ensure that the customers are, in fact, prudently using the retained funds. Self-direct programs vary considerably, but generally allow qualifying customers to self-direct part of their PBF funds to energy efficiency investments in their facilities. Only a handful of self-direct programs are subject to the type of oversight common in traditional industrial energy efficiency programs and thus able to ensure that the retained funds are used in a cost-effective and fair manner (Chittum 2011).

Ensuring Reliable Energy Savings

Ratepayer-funded energy efficiency programs are usually designed to acquire energy efficiency as a system resource. Energy efficiency is consistently cheaper than other types of generation resources (Lazard 2011) and is therefore prioritized as a system resource in a number of states (Friedrich et al. 2009). As a result, 24 U.S. states have some type of long-term energy efficiency target or goal, recognizing energy efficiency's importance as a quickly deployable and highly cost-effective energy resource (ACEEE 2011).

Specific energy efficiency savings mandates or targets often then drive internal utility or PBF organizations' program plans, and the utility or PBF organization's continued access to the collected ratepayer funds is predicated on the assumption that energy efficiency programs are making cost-effective progress towards such goals. Energy system planners, policymakers, and regulators need to know that claimed energy savings within these programs are actually occurring. For energy efficiency to be viewed as a true system resource, all parties must have confidence that savings are real, persistent, and reliable.

There is much at stake, and the budgets for industrial energy efficiency programs continue to increase. In 2010, utilities and PBF organizations reported spending $737 million on industrial energy efficiency programs alone, most of them funded by ratepayers (Chittum and Nowak 2012). That number is considered conservative, and a different report found that total spending on industrial *and* commercial energy efficiency programs in 2010 by electric and gas utilities and PBF organizations was $2,030 million (CEE 2012).

Program Evaluation

To help understand existing energy efficiency programs and guide the development of future energy efficiency programming, programs are routinely subject to evaluation. The evaluation activities are conducted either at the behest of regulators or state agencies, or by the utility or PBF program itself (Kushler et al. 2012). There are several types of energy efficiency program evaluation. These include process evaluation, which determines the efficiency and efficacy of the program itself; market effects evaluation, which determine the longer-term market transformation impacts of a program; and impact evaluation, which determines the actual impacts and benefits of the given energy efficiency program (NAPEE 2007).

This report focuses on the latter type of evaluation, and the following section will describe program impact evaluation components and approaches specific to industrial energy efficiency programs. For the remainder of this report, "evaluation" will refer to *impact* evaluation activities.

What Evaluation Should Reveal

In addition to determining the extent to which energy efficiency programs are meeting goals and general cost-effectiveness criteria, evaluation also determines *how* cost-effective specific programs are. Energy efficiency is often compared to the cost of new generation, in one way or another, and so the true cost of energy efficiency measures must be regularly ascertained to enable meaningful comparisons. Depending on which of the aforementioned cost tests is used, evaluation will show how specific recent or current energy efficiency programs are comparing to the cost of new or existing traditional energy generation resources.

At a minimum, evaluation seeks to determine a verified amount of basic gross savings and affirm that claimed measures were indeed implemented and the engineering analyses used to claim savings look reasonable (Coito 2012). But evaluation can go much further. It often seeks to ascertain the number of free riders, who are participants who earn an incentive for a project they would have done anyway, absent the program. Evaluators may try to calculate non-energy benefits, such as water savings or improved manufacturing productivity. They may also try to measure whether non-program participants have been indirectly influenced to make energy efficiency investments by the program, which is referred to as spillover. All of these issues are deeply nuanced and appear to be almost impossible to measure with total accuracy. However, approximations of these factors can be useful for understanding the merit and value of programs, and also for continuous program improvement.

In practice, program evaluators are trying to unearth some useful facts about energy efficiency programs as a whole by looking at a smaller population and making some educated estimates. This is due to the time and expense of conducting deeply detailed project-by-project or customer-by-customer evaluations to solve for every unknown variable that could impact a project. Instead, evaluation often uses known facts about previous projects or other assumptions to inform the understanding of current or future projects. It is these activities that make evaluation an activity that yields questions about overall accuracy and whether it is realistic or meaningful to seek high levels of precision. Evaluation activities can sometimes feel quite burdensome to program administrators and industrial customers.

Evaluation results can also be quite helpful to program administrators. Many respondents to the online survey indicated that evaluation also helps them get additional feedback on their programs. For example, one program administrator noted that during open-ended questions asked by evaluators, some customers indicated that they wished they'd had someone with more authority than an account manager at the utility to speak to their own CFO. They indicated that having the opportunity to establish a relationship between their CFO and a utility vice president might help further convince their CFO that energy efficiency projects were worthwhile. The evaluator shared that information with the program administrators, who were able to act on the information.

Evaluation activities are typically undertaken by third-party entities that are expected to be unbiased. It is assumed that these firms can rationally determine whether the claimed amount of energy savings is accurate and whether the program itself truly influenced those savings. In addition to providing unbiased assessments of program performance, the inclusion of external parties can help foster the types of serendipitous outcomes mentioned above. By bringing evaluators in to judge certain aspects of a program's success, programs can open themselves up to information that they may not have thought to collect themselves.

Components of Evaluation

The first major element of any evaluation process is development of the baseline. The *baseline* is the facility's energy consumption that would have occurred absent the implementation of the energy efficiency measure, project, or service. The baseline is also referred to as the "business as usual" case (NMR 2010). While the baseline is the energy use prior to the new measure, the baseline is often, as discussed in a future section, developed ex-post facto, or after the new measure has been implemented. This is due to the typical timing of evaluation activities.

After the baseline energy use has been established, evaluators typically measure *gross energy savings*, which are the actual energy savings observed in the facilities that have been exposed to and participated in the energy efficiency program. The gross savings take into account certain elements "beyond the customer's control," such as weather (NAPEE 2007, NMR 2010).

To ascertain gross savings, several methods are used. In a *billing analysis*, evaluators have access to customer energy bills from various time frames and aim to make a conclusion about the change in energy use between the time of program intervention or measure deployment and the post-installation time period. In the industrial sector, a billing analysis will often augment other types of evaluation activities in order to paint a fuller picture. Where possible, bills will be matched to some measure of output to understand pre- and post-retrofit consumption patterns.

Where appropriate, *deemed savings* may be calculated based on pre-determined levels of savings for specific equipment or measures, operating in specific conditions. Some jurisdictions rely heavily on deemed savings calculations, which are based on deemed values that are universally accepted. For some less complex projects, the use of deemed savings may negate the need for direct on-site measurement. Deemed savings are typically used for measures that are used in a fairly consistent manner across a large population of sites. However, industrial projects are typically more complicated and require additional measurement. Kushler et al. 2012 found that 86% of states used deemed savings to some degree, but it appears a smaller percentage of industrial programs regularly use deemed savings estimates in evaluation activities. This is due to the fact that customized evaluation approaches are often required for complex industrial projects involving new equipment, behaviors, and facility operations.

Using *measurement and verification (M&V)* savings estimates, which are often developed by the program, is a common method for determining gross savings in program evaluations. Typically, programs assess the amount of energy a measure is saving, and then use that value to determine an appropriate incentive. Evaluators will often use M&V data to augment their own analyses or as a jumping-off point for their gross savings analyses. As discussed in the next section, the degree to which M&V data is used by evaluators varies widely.

Gross savings are also subject to the application of a *realization rate*, which is the ratio of actual savings (identified during evaluation activities) divided by the tracked or projected savings (typically estimated by program M&V activities). A realization rate of 0.9, for instance, indicates that evaluators believe that internal M&V-based estimates of measure savings tend to overestimate true measure gross savings by an average of 10%. Once gross savings have been adjusted to reflect the realization rate, they become *adjusted gross savings*.

Some programs are only evaluated for gross savings, but many regulators and policymakers want to determine how much of those savings were actually influenced by the energy efficiency program in question. To answer that question, programs are then evaluated for *net energy savings,* which are the energy savings that would not have occurred absent the energy efficiency program (NMR 2010).

To calculate net savings, several other aspects of program impact are measured. The first is an analysis of *free riders,* who are program participants who would have implemented the measures or practices in question absent the energy efficiency program (NAPEE 2007, NMR 2010). The second is an analysis of *spillover*, which are the energy savings that may have occurred when the program, by its mere presence, indirectly influenced measures or behaviors that saved energy, though the customers or facilities implementing those measures or behaviors were not actual program participants. Spillover can also include savings that occur when a program participant implements changes elsewhere in the facility as a result of being a program participant. This is sometimes called *inside spillover*.

To determine free rider and spillover effects, evaluators will typically conduct interviews and surveys among both program participants and non-participants. Questions include why an investment was made and inquiries about the customer's interaction with the actual efficiency program. Surveys of vendors can also be conducted if facilities indicate that the vendors have a substantial influence on project decision-making. According to the online survey of industrial programs, 23% of respondents indicated that a portion of their customers were regularly surveyed over the phone to obtain this kind of information, while 73% of respondents indicated that a combination of survey methods, including via telephone, were used to help measure program influence and determine things like free ridership.

Surveys may be short, ten-question yes/no instruments, but efforts can range to much more detailed interviews of 30 to 45 minutes. These surveys might also attempt to identify additional *non-energy benefits* of a project, which may then be calculated for the program as a whole (Coito 2012, Glanton 2012, Friedman 2012, online survey). Non-energy benefits for industrial customers can include improved productivity, improved worker safety and satisfaction, and reduced water consumption.

A common practice reported among interviewees and online survey participants was to conduct a stratified random sample of projects, but then conduct something closer to a census of the largest projects. The determining factor of what constituted "large" projects varied.

DIFFERENCES IN EVALUATION ADMINISTRATION

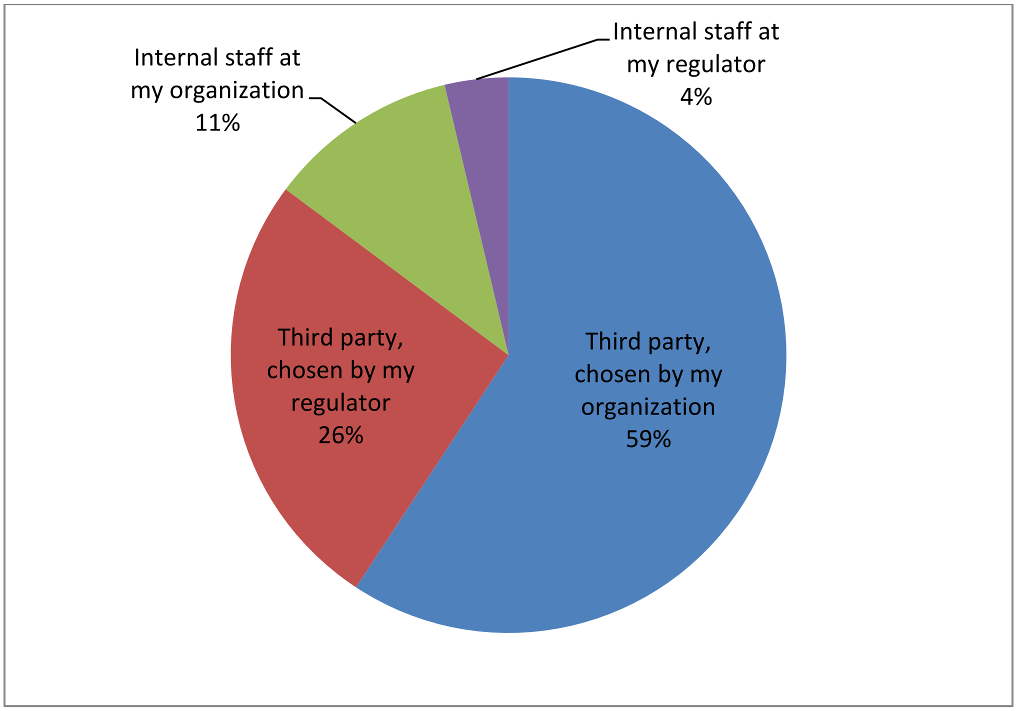
Evaluation looks different from state to state. In a survey of evaluation practices across all sectors in 45 states, a recent ACEEE report found that utilities themselves were responsible for overseeing evaluation activities 36% of the time; utility commissions were responsible 18% of the time; and both utilities and commissions were jointly responsible another 18% of the time (Kushler et al. 2012). According to the same study, 79% of the respondents indicated that external third-party consultants conducted the actual evaluation activities.

Major evaluation activity is largely the domain of third-party consultants, and the use of third-party consultants to conduct evaluations is common among all types of efficiency programs. Respondents to the online survey of industrial programs indicated that consultants were used to conduct primary evaluation activities 85% of the time, typically chosen by the utility or PBF organization itself. Figure 2 shows who is primarily responsible for conducting industrial energy efficiency program impact evaluation activities.

Third-party consultants share best practices and are actively involved in a number of different groups to help maintain the freshness of their approaches and to learn about new techniques and technologies. However, some staff at industrial energy efficiency programs noted that they felt it was critically important to have their staff more involved in the evaluation process to make sure evaluators fully understood the program's internal data and practices. Among industrial programs, there is considerable variation in the degree to which program staff is involved in evaluation activities.

According to the online survey and interviews, regulators are at least somewhat involved in industrial program evaluation activities in most cases. This mirrors what is happening in evaluation across all sectors, where only 9% of states reported that there is no role at all for the regulatory entity in program evaluation (Kushler et al. 2012).

Figure 2. Primary Party Conducting Industrial Efficiency Program Evaluation



Source: Online survey

No one entity is responsible for defining what program evaluation should look like. Program administrators frequently reported that while regulators or policymakers gave "guidance" in evaluation activities, the details were often worked out on a case-by-case basis when evaluators began new rounds of assessment.

Within industrial evaluation, the sheer heterogeneity of the customer class also typically yields very different approaches to evaluation depending on the size of the facility, the type of projects being undertaken, and the degree to which the customer is already clued into his energy use. One user of evaluation results noted that he sees "more variation across specific industries than across different states" (Eckman 2012).

So while there are a wide variety of approaches to evaluation administration, a "typical" general approach does exist. In it, the regulator defines or gives guidance to utilities, who in turn hire a third-party consultant to conduct an impact evaluation within certain timeframes. There is a back-and-forth between the consultants and the program staff during the evaluation process, and the consultants deliver a final report that stipulates the amount of savings for which a program ought to be credited. They also may provide recommendations for improving program savings. Variations abound. Some utilities and PBF organizations have multiple programs evaluated concurrently, and results are thus identified and discussed at the "portfolio" level. Other programs conduct their own mini-evaluations on a nearly per-measure basis, developing their own estimations of free ridership and the like while evaluators only check in every three or four years.

The resources dedicated to evaluation differ as well. Two previous analyses of spending on program evaluation, measurement, and verification (EM&V) found spending ranging from 1-8% of program costs (Kushler et al. 2012, Messenger et al. 2010). In one study, all but one jurisdiction reported spending the majority of their EM&V budget on impact evaluations, which are the evaluations targeted in this report (Messenger et al. 2010). All interviewees reported their own industrial evaluation spending ranges as within these identified ranges, though some industrial programs reported that spending on impact evaluations was roughly equal to spending on other types of evaluation for their programs, including market assessments.

Anecdotally it appears that evaluation budgets are not being decreased and are, in some cases, increasing. Online survey respondents and telephone interviewees generally reported seeing their evaluation budgets kept consistent as a percentage of their program budgets, which are increasing, thus yielding increased evaluation budgets (Chittum and Nowak 2012). Several programs indicated that the actual budget percentage dedicated to evaluation has increased recently as well.

How Measurement and Verification Informs Evaluation

Measurement and verification (M&V) is conducted by efficiency programs to determine the energy saved by specific efficiency measures installed at a certain place, under certain conditions. This differs from evaluation, which typically determines a program's total impact over the course of a whole program period, which may be a year or longer. M&V of energy savings is most often conducted during or right after the process of measure implementation, while evaluation is conducted after an entire program has ended, and often well after a measure has been installed.

M&V actions are often undertaken by program staff themselves and are augmented by the work of the customer or contractors the customer or utility chooses to hire. M&V is used by program administrators to understand the performance of a given measure and assess the degree to which the measure is performing as originally intended or projected. Efficiency programs often rely on their own internal M&V when determining whether a customer has earned a particular incentive, according to the online survey performed for this report. Verification of savings may also continue throughout measure life, to ensure that savings are actually occurring as estimated (FEMP 2008).

M&V activities can form the basis for the initial gross savings reported by programs, or at least offer critical data points, such as whether the measure was actually installed and how it is operating. M&V data is used in many evaluations, and often evaluators have access to program-supplied M&V data when beginning a program evaluation. In particular, according to the online survey, metered data is most often used by and desired by evaluators. About half of respondents to the online survey indicated that any available metered data is reviewed within the impact evaluation process, though the extent to which that data actually enters the evaluation calculations appears to vary considerably.

Though M&V data is not usually the primary data used within program evaluations, evaluators use it frequently. Evaluators compare what they model, see in the field, and learn from customers with the existing program-provided M&V data. M&V data points are often used to supplement evaluation data where holes exist. Evaluators usually conduct their own analysis of energy savings and verify which measures are in place, but if there is high confidence in the M&V data, evaluators may sometimes take the M&V data as their initial set of data points for calculating savings.

Evaluators also noted that they typically review M&V savings calculations to make determinations about the soundness of a program's internal energy savings methodologies. If something strikes evaluators as odd or abnormal, they will likely flag it for further review. Evaluators pay particular attention to the large custom industrial projects, which usually require their own custom M&V models. Evaluators will examine the project calculations and may request the full detailed model from whoever conducted the project M&V. They also ascertain whether internal program estimates of savings were updated to reflect findings unearthed by M&V activities.

Evaluators also use certain elements of M&V to assess the conditions present at the time of project implementation, and confirm assumptions the evaluators have about a project's applicable baseline. Evaluators also use other data points collected during M&V activities, including information about installation rates of certain measures.

In most cases, evaluators use or access only a portion of the data collected from program M&V activities. Only in four instances was M&V data reported to be looked at in total by evaluators. All other online survey respondents indicated that only representative samples of M&V data were considered in addition to other data such as billing information. Increasingly it appears that evaluators are more amenable to using the supplied M&V data as inputs to their evaluation efforts, provided they have faith in the numbers. "If [the M&V] is solid, we'll just go back and collect ex-post data and compare that to the forecast" to evaluate a given project, said one evaluator.

How Evaluation Findings Are Used

Impact evaluations are designed first and foremost to determine the exact energy impacts of a program. Ideally they can also help inform program administrators and designers when planning new programs or considering changes to existing programs.

Regulators use evaluation findings to make determinations about a program's performance. According to Kushler et al. 2012, evaluations of programs across all sectors are almost always used for general program oversight, while 41% of states use evaluation results to determine specific efficiency-related shareholder revenues. Only 3 of the 45 states responding indicated that evaluations were used to determine a program's cost recovery amounts (Kushler et al. 2012). Typically, programs are able to recover their costs regardless of evaluation findings, though poor performance might be remedied or addressed in future program development.

Impact evaluations are viewed by regulators and policymakers as the final word on how well an energy efficiency program is meeting its goals. Utilities often have very specific energy savings goals, for either net or gross energy savings. Unmet goals can result in substantial financial penalties, separate from cost-recovery considerations. In addition to establishing the final assessment of net or gross savings, an evaluation will also often indicate the overall cost-effectiveness of a program, and whether or not a program met its prescribed cost-effectiveness tests. Regulators and policymakers consider these findings when utilities or PBF programs are making their cases for future program funding.

While evaluations are usually conducted for a full program period and used to assess the impact of a program period that is past, the lessons learned from evaluation and internal M&V activities are often used to tweak and alter programs as they are still running (NAPEE 2007). So while there is a defined program period and a point at which the final evaluation report is published, partial findings are sometimes available before the final evaluation report is published. These findings may be used by stakeholders, depending on their needs and the perceived worth of the findings.

Evaluation results are also used by the program administrators themselves to determine how to shape future programming. For instance, if the level of free riders for a particular measure was identified as significantly higher than the prior program period, a program administrator might conclude that she no longer needed to offer incentives for the measure, or at least needed to examine why free ridership was rising. Good evaluations can help program administrators better understand their customers' needs.

Two other types of evaluations can help a program understand if they are meeting the needs of their customer base. *Process evaluations* can determine "how efficiently a program was or is being implemented," and *market effects evaluations* can assess how well a program is influencing the marketplace to encourage "future energy efficiency projects" (NAPEE 2007). As one program administrator noted, process evaluations can also help programs get a much better understanding of whether their contractors are operating in a manner that is well-received by customers and beneficial to the program itself. Market effects evaluations can also feature detailed verbatim responses from customers, attributable to specific projects, allowing program administrators to understand the exact degree to which a customer was satisfied with the program and why.

Evaluation is also used to update deemed savings amounts, which are the universally accepted amounts of energy savings attributable to certain measures in certain conditions. Many programs claim their savings based largely on deemed savings, and so ensuring accurate values for deemed measures is important to program and system planners.

Finally, impact evaluation can play an important role in the development of industrial energy efficiency supply curves. Regulators and system planners engaged in long-term planning of efficiency as a resource understand that the industrial sector is a complex one. They say that evaluation can help identify trends and changes in industrial customer behavior and help accurately forecast industrial energy savings. Evaluation results can be used to update the models used to forecast future energy demand. "We need to understand what the (efficiency) resource is likely to be in the future, and evaluation brings an important key to the table," said one system planner.

The Challenges of Industrial Program Evaluation

The fundamental challenge facing every energy efficiency program evaluation is that it must attempt to estimate a value for something that cannot be directly measured. Specifically, evaluation must estimate the difference between what energy consumption actually occurs and what "would have occurred" in the absence of the program. In essence, it must "measure" energy consumption that is *not* happening—due to the program. By definition, this cannot actually be done, and so instead evaluators try to make the best estimates possible based on available information and time and budget constraints. This fundamental challenge underlies every major operational challenge in industrial program evaluation.

THE NATURE OF INDUSTRIAL FACILITIES

The industrial sector is uniquely challenging for evaluators of energy efficiency programs. A report 15 years ago (Clarke et al. 1996) identify three main reasons. These were:

1. The challenges of developing an accurate baseline,
2. The challenge of accurately annualizing results collected from a shorter time period, and
3. The challenge of normalizing results to post-retrofit service and product production levels (Clarke et al. 1996).

What that report noted is still true today: the industrial sector is complex. It is a highly heterogeneous sector, making comparisons among different program participants very difficult. It is also a sector with production cycles strongly influenced by seasons and economic trends, making post-measure calculations of energy use look very different depending on the month or the year.

"Every site is different, and most require a fair degree of engineering talent to accurately assess projects," said one evaluator. These types of quotes were common throughout the interviews and survey responses, and representative of opinions held by regulators, evaluators, and program managers.

There are tremendously sensitive assumptions that must be made in industrial sector evaluation. An industrial program manager noted that, "in industry, most of your savings are calculated, rather than deemed, measures. It's very hard to presume that even if you are calculating savings at 90% of your facilities," you can use those findings to make assumptions about 100% of your facilities in the future, as is often done. "Maybe something was anomalous? Maybe there was one hot new technology that year that flew under the radar? And yet we use these calculations like they are completely accurate." Energy savings also are typically annualized, though evaluators may only be working with data stretching over two or three months. The changes to and influences on an industrial facilities' energy use over a whole year may not be fully documented by data points from a few months.

Adding to the complexity is the fact that a substantial number of industrial energy efficiency program offerings are custom in nature, meaning that the program is tailored to the unique needs of each customer. Rather than offering incentives for particular activities or technologies, custom energy efficiency programs reward customers for making any number of energy efficiency investments or even behavioral changes. "It's never a one-size-fits-all" approach to evaluating these programs, said a program manager. Each custom project often requires its own evaluation approach, whereas other sectors can use a single evaluation approach to assess an entire program.

While industrial projects are often subject to custom internal evaluation, measurement, and verification of savings within the given program, it is the evaluation approach deemed appropriate by the regulator that usually prevails. This may or may not be a method or approach used by program staff. Thus, program managers and regulators both noted the all-too-frequent occurrence of many hours spent evaluating a project internally by a program, only to be reassessed by an evaluator at the behest of a regulator later on. "This wastes time, it wastes money, and it wastes resources" to not have a predefined approach to program evaluation, said one policymaker. Internal program evaluation findings could be used by external evaluators if a clear protocol existed and was followed. Some industrial energy efficiency stakeholders suggested that a unified and standardized approach to evaluation would reduce redundancies and help improve the transparency and predictability of external evaluation activities.

Programs and evaluators are also challenged by data collection issues. Industrial facilities are often loath to reveal energy use, production line operations, and equipment data that they consider to be proprietary. Gathering data necessary to verify existing M&V data and performing additional onsite measurements can be highly burdensome to a customer. Customers complain to program managers about the inconvenience, but, as several evaluators noted, there will always be some level of inconvenience to the customer compared to doing nothing. Site access was a difficulty expressed by a majority of evaluators, who noted that it was not unusual to be missing one or two major projects in a given sample due to uncooperative customers who made access difficult.

This lack of data can yield reduced confidence that programs are being evaluated with reasonable accuracy, given the purposes of evaluation. As noted below, animosities and distrust among some evaluation stakeholders is not uncommon, and with less "good" data, more assumptions must be made that can cause consternation among all parties.

EVALUATION'S PLAYERS

There were many interviewees and online survey respondents who appeared to view program evaluation as a necessary evil—that is, something that is clearly needed and serves an important purpose, but something that could be greatly improved. The opinions varied dramatically depending upon the role played within the evaluation process.

Among program administrators and managers, a main concern was the fact that they felt their programs' nuances were not appreciated by evaluators and thus not understood by regulators. They believed their unique relationships with their customers were not properly accounted for and credited in evaluation activities. Program managers also felt that evaluators entered into evaluation activities with a feeling of superiority and a general belief that the data collected from the customer was going to be better or more accurate than what was collected by the program. Additionally, program administrators noted that evaluators frequently extrapolated from a small sample to make determinations about a much larger population in a manner that seemed inaccurate to them.

Among evaluators, frequent concerns included a general sense that program administrators were not collecting or providing the quality of data they needed for their evaluations. They also lamented that programs did not always keep their project files organized and their contact information updated.

Evaluators noted that they often have to spend substantial time digging around for data that someone on the utility or PBF program staff had collected at one point, but had not filed properly or in an easily accessible manner. An improvement in data collection systems would speed up the evaluation process, the evaluators said. They also noted a general reluctance from utilities when evaluators asked them to return to a customer to gather more data, for fear of overburdening the customer.

To be sure, not every program administrator spoke negatively about their evaluators, and most evaluators were neutral in their overall assessments of the programs they worked with. In instances where evaluators have a long and fairly harmonious relationship with the staff of the programs they evaluate, it is clear that the evaluation findings are viewed as more useful by program managers and designers. "Two-way communication is critical," said one program manager. "Remove that, and the evaluation results wouldn't be reflective of what actually happened."

The interviews and surveys pointed to a need to improve communication between program administrators and evaluators in general, and to better address the unique aspects of the industrial sector. The next section will discuss specific aspects of industrial energy efficiency program evaluation that could be improved. Such improvements could likely help ease some of the tension between program administrators and evaluators, and yield evaluations that are more useful and responsive to the industrial sector as a whole.

Six Major Issues in Industrial Program Evaluation

Evaluators are most often trying to tell a story about an entire program, or suite of programs, by using representative data from only a portion of the participants in these programs. The logistics and costs of conducting such detailed evaluation of each and every program participant would be highly difficult to overcome and justify. However, while other sectors lend themselves quite well to sample analysis due to their homogeneity, the industrial sector is a difficult one for which to make generalizations.

Every customer is different and many projects are unique. Despites this, some assumptions must be made in the evaluation process in order to provide evaluations that are not prohibitively expensive or unduly burdensome to all involved. These assumptions are often what underlie some of the tensions and challenges of industrial program evaluation.

There are six major data points and practices in industrial evaluation that interviewees and survey respondents coalesced around as most pressing to the practice of industrial energy efficiency program evaluation. They are:

- The development of a baseline energy use estimate
- The timing of evaluation activities
- The use of net savings estimates and net-to-gross ratios
- The measurement of free ridership
- The measurement of spillover effects
- The measurement of non-energy benefits

The following section will discuss the above six issues in detail and, where possible, identify best practices and possible paths forward. What is very clear from the survey and interviews conducted for this report is that few best practices currently exist in the area of industrial energy efficiency impact evaluation.

Developing a Baseline

The baseline energy use of a facility is what the facility *would have consumed* absent the energy efficiency program in question. It is thus not an easily measurable quantity. While the baseline could be measured at the beginning of the program activities, evaluation activities often don't begin until well after a program has started or even ended. Barring time travel, an evaluator cannot actually physically measure a facility's baseline energy use when an evaluation begins at any point after the program has begun to influence the customer.

Complicating the challenge of developing an accurate and reliable baseline is the fact that over the course of the months or years of an energy efficiency program, an industrial facility might experience seasonal and year-to-year peaks and dips in operations, additions of new product lines, reductions of product lines, and changes in staff and equipment that impact the facility's energy use. So while new investments in energy-efficient equipment may have reduced a facility's energy consumption, the facility may also have dropped a product line. Just measuring the before and after energy use of a facility will not tell the full story.

As one evaluator described it, the baseline of an industrial facility could be the:

> Least efficient, non-regressive, code or regulations-compliant option specific to a particular facility and application that the customer technically, functionally, and economically could have alternatively considered to deliver the post-retrofit level of production or service (Maxwell et al. 2011)

A major problem in the industrial sector is that there are no codes and standards to which a facility is subject to on which to base the baseline. There are no standard minimum efficiencies on most pieces of equipment inside a facility. The emphasis in developing a baseline is not what the customer would have done, but more what they feasibly could have done. It is impossible to know what a customer would have chosen to do, and the question of what a customer *would* have done is more appropriate in the later analysis of free ridership (Maxwell et al. 2011).

While the program administrator or customer often develops baseline estimates, evaluators typically make determinations about the accuracy and legitimacy of the baseline. If an evaluator believes that a baseline was set too high, the final gross or net energy savings attributable to the program may decrease when the baseline is reset lower.

Stakeholder Perspectives

There is general agreement among everyone involved in industrial evaluation that developing an accurate and widely accepted baseline is the most important and difficult element of evaluation, as it is the "before" against which all of the "after" is judged. The added degree to which industrial projects are unique and customized requires an even greater scrutiny of the baseline than in other sectors.

The biggest challenge to developing baselines that are viewed as fair and reasonable is, according to program managers, the practice of developing baselines ex-post-facto. This is due largely to the timing of evaluation activities (discussed in the next section), but also with the challenge that so much of the baseline is hypothetical.

"Setting the baseline is, to some degree, the responsibility of the program," noted one evaluator, saying that if program administrators could develop baselines that were satisfactory to evaluators, they might be able to work better together during the rest of the evaluation process. The baseline is the first point at which program administrators and evaluators appear to diverge, and the relationship between the two can be problematic if the program administrator does not believe the evaluator's estimate of baseline out of hand.

Evaluators indicated that they felt it was critical to get agreement on the baseline as early as possible. Many programs do not actually emphasize the development of a baseline, and so while initial measurements of energy use are taken before measures are implemented, the context within which the measurement was done may not be documented. With some evaluation activity happening several years after a project was completed, identifying the exact conditions in which the original measurement was taken can be nearly impossible. A defined process that encourages program managers to document major events and changes that could impact that baseline could help ensure that all parties are working with the same information come evaluation time.

In Practice

In New York, the New York State Research and Development Authority (NYSERDA) staff tries to take a more active role in the development of baselines than many other programs. Data holes that emerge during the development of the baseline can be filled by NYSERDA staff, which develops their own baseline as part of their normal reporting process (Glanton 2012).

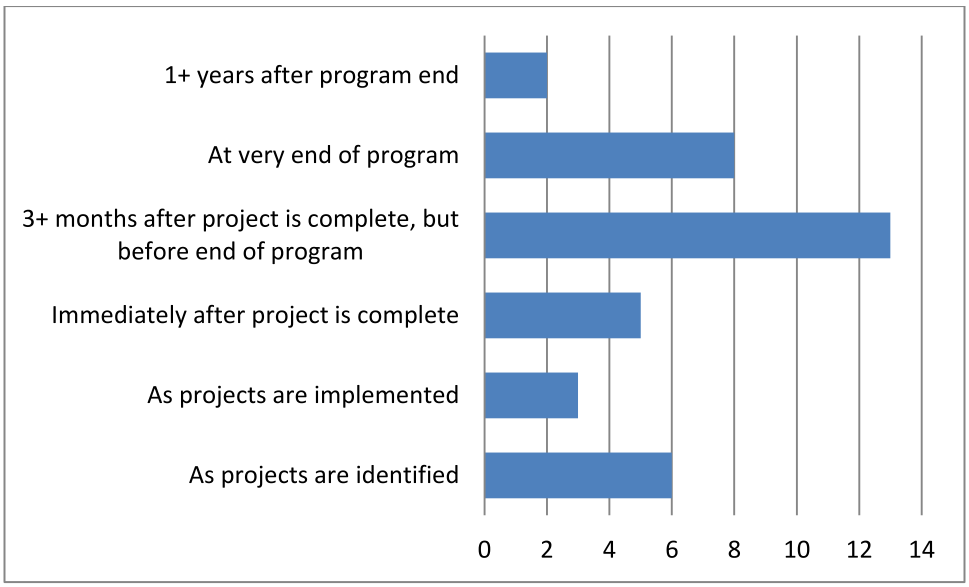
NYSERDA program staff and contractors work to describe the existing conditions in a given site, conducting pre-site inspections that yield descriptions of all equipment present, how it is used, and relevant operating conditions. The baseline is explicitly called out and the elements of the baseline are explicitly described, so there are no questions about what should and should not be included at evaluation time. Under a new pilot, members of the evaluation team are also involved in the earlier baseline development. This increased clarity has been well-received by the program administrators.

However, NYSERDA is still faced with the challenge of determining the part of the baseline reflective of what a customer *could* have done. Additional production lines and expansions of facility capacity can be difficult to reflect in the baseline and future evaluation assessments in a manner satisfactory to all parties.

TIMING OF EVALUATION ACTIVITIES

While programs may operate over multi-year periods, evaluation activities are not necessarily conducted concurrently with the programs. Figure 3 illustrates when industrial programs typically see major evaluation activities. The most frequently noted timing was three or more months after project completion, but before program period end. Many program administrators indicated their evaluation activities were conducted at the very end of the program period, or over one year after the program period had ended. This indicates that a substantial amount of evaluation activities are happening after, or well after, the measures have been implemented.

Figure 3. Timing of Major Industrial Efficiency Program Evaluation Activities (Number of Reported Incidences)



Note: Respondents could indicate multiple major time periods of evaluation activity
Source: Online survey

The type of evaluation activity conducted at various points of the program cycle is important. An energy realization analysis, which measures actual savings versus estimated savings, could be conducted too early, before the usage patterns of a measure are fully known. An analysis of a customer's decision-making behaviors conducted too late might be flawed because a customer cannot accurately recall his or her reasoning or influences.

Stakeholder Perspectives

According to one major evaluation guide, "evaluations should be done within a program cycle so that feedback is frequent and systematic and benefits the existing program and informs the design of future programs and their evaluation" (NAPEE 2007). In practice, evaluation is rarely done entirely within the program cycle, and is often done at irregular intervals that do not coincide with program periods.

Evaluators are generally coming into a project or facility well after the initial energy use has been measured and the decision to actually make the investment has been made. Since those are two major areas of inquiry for the evaluator, he or she must rely on interviews and surveys and secondary data to make determinations about the facility's past and current performance.

Surveys conducted well after the investment decision was made are particularly challenged to reflect decision-making activities accurately. As one regulator noted:

> Few if any evaluators have ever worked in a large industry where several decisions similar in scope and cost to the one to move forward with an energy efficiency project are made weekly. Asking a busy manager the influences [that affected] their decision on any one project, three months after the decision was made, is like asking you where you filled your car with gas three months ago and why you decided to stop at that particular gas station. Most energy efficiency project decisions are far less significant and memorable to the industry than the evaluators realize.

A major trend recognized in the survey was the increase in the number of programs that see faster turnaround time of evaluation activities and findings to help address these challenges. "It's been the biggest difference in our programming recently" said one program manager about the fact that initial surveys of customers about their decision-making processes are happening a year earlier than in prior program periods. According to the online survey, two industrial programs that recently began conducting free rider surveys much earlier than in previous periods found that their free ridership levels dropped very significantly.

"We have found that with large industrial projects the evaluators often have issues with baselines as an upgrade or new construction. It's better to handle that up front," said one respondent on the online survey. By bringing evaluators into the project at the same time program staff begin their involvement, some of the major points of contention appear to have been somewhat reduced.

Evaluation activities happening concurrently with program delivery can also mean that programs can immediately address issues that may have previously only been revealed a year or more after a program ended. "A good evaluation should continually give you good information. If you see that something needs to change, or a baseline needs to be updated, or a process could be improved, you can act on that information now," said one program manager who has seen evaluation activities begun much earlier in the program cycle. Evaluators could continue to be objective observers, but could also provide a helpful service by identifying issues as the programs are being run.

In Practice

The Energy Trust of Oregon (ETO)'s Fast Feedback program is a new, expedited phone survey that occurs 1-2 months after incentives are paid on projects. Whereas the same survey efforts used to be done 1-2 years after a project happened, Fast Feedback can promptly connect with individuals who are familiar with the project. Fast Feedback also is a very short (less than ten question) survey tailored specifically to the needs of each ETO program.

"In prior years, we'd call someone a year or so down the road and expect him to be able to recall exactly why he did a project. Now our participants are being asked to recall details of their decision-making much earlier. In this way, we have much more reliable numbers when we're looking for questions of influence," said Kim Crossman, ETO's industrial program manager.

Since the implementation of Fast Feedback, ETO has seen the rates of free ridership on their industrial programs drop dramatically, greatly enhancing the programs' overall cost-effectiveness. The critical lesson learned was that disentangling the analysis of technical realization from that of the program's influence is very important. Those two aspects of evaluation do not move on the same time line, and seeking answers to both at the same time was not serving the program well. Questions about the process and influence of the program itself are separated from longer term assessments of the program's overall impact, which take a longer time to be fully realized.

The Fast Feedback program was specifically designed to reduce the burden on respondents (ETO 2011). It has also allowed programs access to real-time feedback about their programs that can be addressed much faster than in previous program periods.

In California, a new emphasis on ex-ante review is bringing evaluators into project sites in conjunction with utilities' initial visits. The California Public Utility Commission (CPUC) and the evaluators they hire are now more involved in estimating individual project impacts and rates of free ridership toward the beginning of some select projects. This approach is supposed to help guide utilities and contractors and help them understand "what is and is not flying" prior to the end of the program (Coito 2012).

Some utilities expressed interest in having early involvement by the CPUC, but have been frustrated by the high levels of free ridership identified in these early assessments, due largely to a broad definition of what constitutes standard industry practice. The early analysis has improved the transparency of the CPUC's decision-making, but has caused the net savings of some projects to be significantly discounted from the beginning, as well as during post-program evaluation.

Another challenge of this increased evaluation activity is that industrial facilities tend to be resistant to constantly opening their doors to evaluators at multiple periods throughout the project. A number of programs, in addition to those in California, noted that increased early evaluation activities can sometimes "choke off" projects when customers are inconvenienced by the amount of on-site analysis that must be conducted. In some cases, delays in completing required evaluation steps have led to delayed projects, leaving customers frustrated and considering cancellation of the project.

Net Savings and Net-to-Gross Ratios

Net savings are the savings a program can claim to have influenced. A net-to-gross ratio is the applied ratio of net savings versus total (gross) savings reported. Programs subject to stipulated net-to-gross ratios will see their gross savings multiplied by a net-to-gross ratio, which is often derived from previous evaluations. This calculation will yield net savings.

These metrics are important for ratepayer-funded energy efficiency programs. Those who contribute to these funds want to know that their dollars are being used in a worthwhile manner, ideally to encourage energy efficiency that would not have otherwise been acquired. However, the measurement of net savings is an imprecise exercise that some energy efficiency program managers and policy makers argue is a poor use of resources.

While all programs are evaluated for gross savings, not all are evaluated for net savings. Across programs active in all sectors, 29% of states reported that they did not measure net savings in their program evaluations (Kushler et al. 2012). Continuing that trend, 27% of respondents to the online survey indicated that their industrial programs were not evaluated for net savings. A full 64% of online survey respondents indicated that their programs are evaluated for net savings. The remaining 9% indicated that some industrial programs are evaluated for net savings, and some are not.

Net savings is derived from gross savings. It reflects the overall program savings, after certain additional effects are calculated and netted out. The most frequent additional data points used as inputs when deriving net savings from gross savings are free riders and spillover. Free rider analyses are concerned primarily with questions of influence, which is to say to an industrial customer, 'to what extent did the program itself help you decide to make an energy efficiency investment?' The answer to that question is quite nuanced and complex, as industrial customers are exposed to a wide variety of influences every day. Spillover measures the impact of a program on non-participants, though non-participants are not evaluated to the extent of participants. Those issues will specifically be addressed in the following sections.

While certain programs are given a new net-to-gross ratio within each evaluation cycle, many evaluations simply use a previously derived net-to-gross ratio and then apply it within the current evaluation. Six programs in the online survey shared the net-to-gross ratio typically applied to their industrial programs. Table 1 lists the six ratios reported by online survey participants, showing a wide variation in reported ratios. The gross savings of these respondents are reduced by a widely varying amount depending on the applicable ratio. According to another source of information on impact evaluation practices, the typical "stipulated" net-to-gross ratio is 80% to 95% (NAPEE 2007)

Table 1. Reported Applied Net-to-Gross (NTG) Ratio

Respondent	NTG Ratio
A	1.00
B	0.94
C	0.90
D	0.90
E	0.80
F	0.50

Source: Online survey

Stakeholder Perspectives

As it is necessary to determine whether energy efficiency programs are spending their public and ratepayer dollars prudently, quantifying net savings, and trying to determine how much of the gross savings were actually influenced by the program is one way to attach a value to the work being done by the program. One argument for net is that by measuring net savings, ratepayers will be protected from paying for energy efficiency measures at facilities just trying to "game" the system, by taking incentives for projects they would have done anyway, absent the presence of the program.

Evaluating for net savings can also help program designers ascertain how a program is working, provided they have faith in the findings. One theme from the online survey is that program managers do not necessarily believe what their evaluations claim. This has the potential to reduce the overall impact and influence of the net-to-gross findings. This is troubling because programs are tasked with reducing future load growth and need to know the savings they have influenced.

Evaluating just for gross savings leaves some details unmeasured. A utility or PBF organization account representative could hear about a project being considered at a customer site and then work to offer an incentive for the project, even though the project would have happened anyway. The utility could report higher savings and the customer might receive some level of customer service, but the ratepayers may not be truly enjoying additional efficiency.

The question is whether the threats and concerns that justify evaluating net savings are worth the added cost and effort to ascertain it. This is especially true in the context of major changes to industrial energy efficiency practices, where resource acquisition and market transformation goals are becoming more intertwined.

The arguments against net savings are different from the arguments against the current *approaches* to net savings. The arguments against calculating net savings itself include the fact that substantial resources are used to calculate something that may not be of high value to calculate, when other, less expensive measures of performance might be sufficient. Another argument expressed by interview and survey respondents, echoing earlier research on opinions in this area (NMR 2010, Kushler et al. 2012), was that net savings is not actually something that can be precisely calculated. These people said that gross savings and net-to-gross ratios are rough estimates, at best, but are treated by regulator and policy makers as precise numbers. They felt that these numbers were often not precise or reliable enough to be used for policy and program decision-making purposes, and yet that is exactly what is done.

"It's a perception reinforced by lots of interested parties that it's possible to actually identify energy savings with a high level of precision, and net out what would have happened anyway," noted one evaluation expert. An analysis of the opinions on calculating net savings in the Northeastern U.S. found that net-to-gross analyses create "the impression that the estimates are accurate," despite the fact that deriving net savings may be a matter of piecing together multiple estimates and other less accurate data points (NMR 2010).

Less precise assessments of efficiency program impacts have a place in program and resource planning. They can still be useful to get a general sense of whether a program is performing well, and whether it is truly helping to acquire a certain amount of efficiency resources. However, policy makers and regulators need to understand that such numbers are *not* the perfectly precise numbers that they might desire, and thus need to be reported and discussed with all caveats fully explicit.

In one case, an online survey respondent reported that rather than have a specific program for which a net to gross ratio is determined, each measure is evaluated across all customer classes. The net-to-gross ratio is then determined for each measure specifically. After each measure is assigned a net-to-gross ratio, all the measures in projects implemented in the industrial sector are aggregated together to form a proxy net-to-gross ratio for "industrial" offerings. Such assessments of net savings are likely not truly representative of industrial programs as a whole, since measures are done for a variety of reasons and influenced by many different actors, especially when looking across multiple customer classes and over a wide variety of custom projects.

"Fighting over money leads to the need for precision even when it's impossible," said one internal program evaluator. Indeed, programs that are evaluated to determine shareholder incentives or other performance-based revenues such as lost-revenue adjustments seem most tied to high levels of net-to-gross precision. The concern with an unnecessary and unachievable level of precision was the most common comment about net savings during all the interviews and surveys. "We're using evaluation to determine shareholder value down to the last penny when what we really need to know is, how are we doing [generally]?" said one program manager.

What makes the finding of net savings so difficult is that a large part of the estimation involves determining what would have happened absent a program. These estimations cannot be fully proven, because there is no true "without program" scenario to offer empirical evidence. Measuring free ridership is the primary way programs are assessed for what their participants would have done without the program. There are many concerns with how free riders are determined, which are discussed in depth in the next section. The use of free rider analysis, though, was viewed by about a quarter of interviewees and respondents as a waste of resources. These people and others felt that the heavy attention to free riders came at the expense of measuring other program impacts that are currently not heavily studied, such as non-energy benefits.

The focus on net-to-gross savings also does not mesh well with some of the newer energy management and energy performance programs found around the country. Since some of these programs are both resource acquisition and market transformation programs, they are explicitly aiming to acquire energy resources now and influence the market to make long-term and sustained change in the future. The long-term changes might influence other actors, such as equipment vendors, who, down the line, influence new efficiency investments in other facilities. While such activity would be a success from a market transformation perspective, if a customer cited the influence of a vendor as the reasoning behind an efficiency investment, the original program would likely see those savings netted out of their gross savings. Net-to-gross assessments may not be able to adequately address the nuances of energy management and energy performance programs. "It's so hard to tell whether you're doing a good job measuring net," agreed one program evaluator.

What did emerge throughout the interviews and surveys was a sense that a middle-ground approach to evaluating program impact might serve the needs of regulators and policy makers while providing useful feedback to program planners and designers. "You need to get rid of net-to-gross," said one program manager. "It's a really bad and distracting metric." Instead, said this person, determine a way to derive a reasonable estimate of how well the program is doing.

"I know, anecdotally, that about 80% of my program is running really well, while about 5% needs substantial improvement," said another program manager. "The question is, which 5% is the worst-performing? Which 5% is wasting ratepayer dollars? And what can be done about it? That is what I want evaluation to tell me." This same person indicated that evaluation results are meaningless to him because he does not agree with nor believe in the methodology.

One approach suggested by several respondents was to measure net savings, but measure it less frequently, focusing primarily on deriving gross savings on a regular basis. Evaluations to determine net savings could be done for the most recent year every four years or so, to answer the question: is most of this portfolio being acquired in a prudent manner? Regulators and program administrators could determine a range or threshold for net-to-gross they were comfortable with, and then have the program evaluated for satisfaction of those set parameters.

Another approach is to single out specific measures, such as those representing the largest amount of claimed savings, and subject those to net-to-gross analysis while leaving the measures representing the smaller portions of savings alone. This is currently being done in California, where "high impact measures"—that is, those representing more than 1% of claimed savings—are assessed for net-to-gross ratios (Friedman 2012).

The most popular suggested approach, however, was to move to a more qualitative assessment of program impacts. The widespread perception by program managers that the effort put into finding net savings is not justified by the usefulness of the final number was paired with a sense that finding the impact of a specific program vs. what would have happened absent the program was still important. Determining gross savings, and then determining which projects were and were not *likely* influenced by the program is useful information. "We don't need the answer down to more than one significant figure," said one program manager. Regulators might question the high degrees of precision reflected in net savings estimates. However, having a sense of the rough percentage of projects that were not likely influenced by the program could help significantly in program planning and in determining utility incentives, so long as all parties agree about the need to use such approximations.

In Practice

In the public utility system in the Pacific Northwest, where efficiency is slated to meet over 85% of the area's load growth over the next 20 years, system planners have little use for net-to-gross calculations (Eckman 2011, NWPCC 2010a). Net savings is viewed as a less useful metric than gross savings for purposes of long-term energy resource planning (Kushler et al. 2012, Gordon 2012, Crossman 2012, Eckman 2012). Energy planning for the region relies on real (gross) savings to materialize, which is what really impacts resource planning models. Planners just need to know that measures were put in place, the amount those measures are saving, and how long those savings will last. Whether a particular program did or did not influence the savings is immaterial to resource planners.[3]

However, the utilities and PBF organizations in the Northwest do estimate net savings. While the regional planners are tasked with forecasting an amount of conservation, regardless of the method by which it is achieved, utilities and PBF organizations are accountable for an amount of conservation their programs add, in addition to the conservation provided by the market. The emphasis on precision in calculating net savings, however, is more relaxed than in other areas of the country.

As Kushler et al. suggest, the determination of whether to seek net or gross savings should be made by those who understand what the needs of a given utility territory, state, or region are (Kushler et al. 2012). The Northwest Power and Conservation Council (NWPCC) develops and updates twenty-year conservation and electric power plans for the Pacific Northwest, which includes Washington, Oregon, Idaho, and Montana. The portions of these states served by public utilities and, to a lesser extent, investor-owned utilities, rely heavily on the hydroelectric power supplied by the Columbia and Snake Rivers and marketed by the Bonneville Power Administration (BPA). The power plans, updated every five years, serve to offer long-term projections of future energy supplies and demands for the area (NWPCC 2010a).

Though the NWPCC is not an official regulator, it is tasked by law to develop plans for BPA to meet the region's electrical needs "at the lowest possible cost." It is also required to give priority in its plans to "cost-effective conservation," ahead of renewable energy resources (NWPCC 2010b). Since BPA supplies about 40% of the region's power, the goals and approaches of the NWPCC tend to heavily influence those of the public utility commissions, utilities, and PBF organizations in the region.

ETO conducts evaluations of its industrial programs on a fairly infrequent basis (e.g., every 2-3 years), encompassing all activity within that period. The timing is designed to balance the need for currency of results with labor loads and cost. This schedule is possible because ETO does not earn regulatory incentives on efficiency, so the primary evaluation need is to stay current enough to manage programs well. ETO industrial program realization rates are relatively stable, so waiting so long between evaluations is viewed as low-risk.

In the impact evaluation, the evaluator reviews assumptions made by ETO staff and contractors, and determines whether the assumptions are reasonable. Claimed savings are tied directly to specific measures the ETO is aware of, and any savings that cannot be directly linked to a specific measure is not included in the total savings amount (Crossman 2012). Evaluations use site visits, logs, billing data, spot metering and revised models as appropriate.

---

[3] It should be noted, however, that recent requirements that Washington State utilities acquire "all cost-effective" energy efficiency might change the manner in which regulators and policy makers in that state view the calculation of net savings.

As one of the ETO's internal evaluators noted, the ETO has no profit incentive to claim savings, and thus does not have contested fights with its regulator about net savings. Evaluations are subject to multiple levels of independent review, and ETO has established conventions with their overseers for how market effects and impacts are measured. They have found an approach that works for them and they "live with the results" (Gordon 2012). Resources that other utilities and jurisdictions spend on more expensive—but not necessarily more precise—methods, and on contesting and defending results, can instead be dedicated to other evaluation activities, such as process evaluations, which help the ETO identify ways to improve their programs.

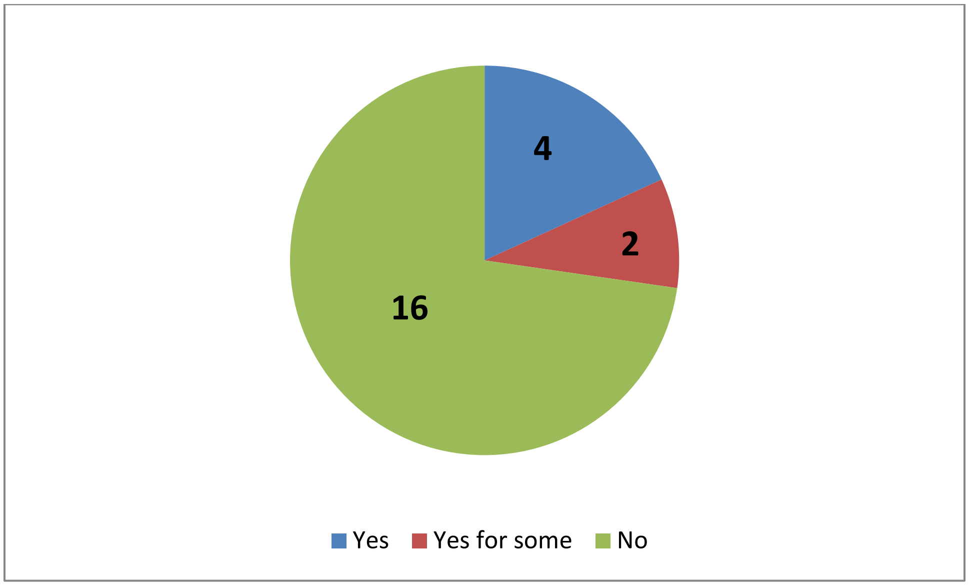
FREE RIDERS

Free riders are those that participate in a program, such as by earning an incentive, but would have implemented the measure or practice in question absent the program. The definition of free ridership, the manner in which it is surveyed for, and the use of default values in measuring net savings are the three biggest concerns respondents voiced about free ridership. In general, respondents also questioned the general value of determining free ridership.

Free riders are measured because regulators, policy makers and program managers want to know where program dollars are being spent on acquiring efficiency resources that would have been developed anyway. The idea is that money spent on free riders is wasted or a poor use of funds. Since the efficiency investments would have happened regardless, the money spent on free riders could have been spent on other customers or services. Free riders, in theory, represent efficiency that happened to occur concurrently with a program's activities, but not due to them.

Free riders are a common data point to collect about energy efficiency programs. For programs operating in all sectors, 67% of states reported that savings results were adjusted for free riders (Kushler et al.2012). About the same percentage of industrial programs reported that their programs were adjusted for free riders. Most of those use established multipliers to determine free ridership, and, according to the online survey, most of those multipliers are based on earlier evaluation results of related programs. Among industrial programs, 27% of respondents indicated that a default value is used to calculate free ridership in all or some of their industrial programs. See Figure 4 for the full breakdown of respondent answers. If not a default value, online survey respondents indicated that their applicable free rider rate is usually determined at the time of evaluation.

Figure 4. Number of Respondents Indicating Whether a Default Free-Rider Value is Used



Source: Online survey

So while most industrial energy efficiency programs are not subject to an automatic multiplier to determine free ridership, most are measured for free ridership at some point within the evaluation process. On the other side of the coin of free ridership is spillover, discussed and defined in the next section. Though tremendous effort is put toward ascertaining free riders, very minimal effort is currently put toward measuring spillover effects. The concerns about free ridership are exacerbated by the fact that a countervailing analysis to that of free riders is not usually conducted.

Stakeholder Perspectives

While many people have noted that accurately assessing a counterfactual—that is, what would have happened without the program—is impossible, because there is no "without program" situation to observe, there is also the added complication that for most industrial firms, many things influence a decision to invest in energy efficiency measures (Blumstein 2009, NMR 2010, McKane 2012).

There is a substantial disconnect in the way free ridership is typically defined and the manner in which industrial companies decide to make energy efficiency investments. Over the course of months and years and decades, manufacturing firms are influenced not only by their utility or PBF program, but by their competitors, companies up and down their supply chains, their equipment vendors, their customers, and individuals at conferences or trade shows. "What they would have done anyway is a pretty soft science," said one evaluation expert. This person, and other interviewees, believed that identifying free riders is far easier for residential customers than in complicated and complex organizations such as manufacturing firms.

Free rider numbers are often not trusted by program managers because major aspects of free ridership are effectively self-reported. One evaluator explained, "Individuals lose track. They might ask themselves: where did I get this idea from? There's a general intelligence of the industrial sector that goes up, as people talk and do business, but it's hard to pinpoint where these ideas originated." Evaluators use surveys and interviews to glean information about the decision-making activities of a firm, and ask where an idea for an investment came from and whether the utility or PBF program had a role or influence. Of the programs that do measure free ridership, 96% of them reported that a telephone survey or another type of survey were the primary method used to determine free ridership.

Most program administrators interviewed believed that the use of surveys, especially one or two years after the project was done, is not an accurate way to determine free ridership. "Respondents think they need to be good corporate citizens by saying that, yes, they would have done that project anyway," said one program manager. People may believe, retrospectively, that they would have made an energy efficient investment, because they like to think of themselves as the type of person or the type of business that would be making such responsible decisions. However, while a utility or PBF program may have been the final reason why a project happened, the individual surveyed may state that they would have done it anyway, even if it truly would not have ever happened (NMR 2010).

There are certainly some true free riders in any efficiency program, but even the definition of free rider causes confusion and frustration. For instance, if a facility would have made an investment anyway, an energy efficiency program may have encouraged that investment to happen sooner rather than later. "It's a grey area," said one evaluator. "Any company not hunting for money is doing their shareholders a disservice." A program evaluator using the standard definition of free rider would likely find such a company to be a free rider, though the investment may have been hastened by the general knowledge that there are incentives available for efficiency projects.

Some evaluators and regulators stated that the most important reason to measure free riders was to identify and address participants who might be "gaming the system" by earning incentives on projects they would have done anyway. The idea is that programs should not be able to claim savings associated with these types of participants. While there are indeed industrial customers that are savvy enough to plan for and undertake projects and then seek incentives, program managers stated that they believe that threat to be low. Instead, most program managers noted that there are far more players in the market who are making energy efficiency investments without incentives or support services of their energy efficiency programs—an effect called spillover, discussed in detail in the next section.

The issue of free riders is even murkier when industrial energy efficiency programs engage in market transformation activities, or activities designed to both acquire efficiency resources and encourage long-term changes in the efficiency marketplace. Arguably, any program that has been around for a long period of time and cultivated and maintained a wide range of happy customers has engaged in long-term market transformation, by introducing firms to their high performing peers and to new technologies and practices and behaviors. They have also influenced third parties and outside players, such as vendors. Customers may originally learn about an efficiency opportunity from a program activity and not implement it for years. Or they may learn about it from a vendor that only has begun offering the technology because of the program's existence. Once the project is implemented, though, the original impetus—which could have been the energy efficiency program—has been muddied in new facility needs or vendor influences. The program was still instrumental in causing the development of that efficiency resource but likely will not get credit within an evaluation. An improved methodology for determining program influence that includes vendors would help overcome this issue.

Finally, free ridership would likely not be such a big frustration to program managers if the free rider value was not so commonly used by regulators and policy makers to judge program performance such as net savings and determine whether certain incentives should be discontinued. Higher free rider rates may indicate that, in fact, enough of the measure or practice has infiltrated a sub-sector that it has become common business practice to do something. It may be common practice, though, because a certain activity or technology is encouraged by an energy efficiency program. So while free riders might be higher, so too is spillover, since the market has picked up on the new technology or practice just by sheer popularity within a sector or geographic area. While at some point it might be prudent to stop offering the related incentive or service, it may be that allowing the program to run with a high level of free ridership is actually serving larger market transformation goals better than discontinuing the program would. Similarly, a free rider rate of zero might indicate that the program has not successfully impacted the market in any significant and lasting way.

“Once we’re successful [at market transformation], we stop being able to take credit for our success, because regulators get concerned about higher levels of free ridership. But should we stop our successful efforts?” asked one program manager. A high free rider value can make a program look less attractive to regulators and policy makers, and the utility or PBF program may then have difficulty justifying its continued existence. Free ridership might be concentrated in one particular subsector or for one particular measure, but a program is often assessed for free ridership in the aggregate.

High levels of free ridership can also impact net savings in the future, because default free rider rates are influenced by prior evaluations. Even if additional information changes free ridership estimates, Kushler et al. report that changes to free rider estimates are typically only made for future evaluations. So if an evaluation finds that free rider rates are too high (or too low), that finding will likely not affect overall free rider numbers until the following evaluation (Kushler et al. 2012). This can mean that fully understanding the impact of a current program is obscured by using dated free rider rates.

New types of industrial energy efficiency programming may render free ridership a moot point. For programs that offer on-bill financing and minimal incentives, program administrators believe the use of the program itself is indication enough that a customer is not a free rider. “If they had to borrow 90% of the project cost, if they were willing to take on debt, surely they must have needed the money to do it. In our view, there cannot be free ridership [among users of our on-bill financing program]” said one program administrator.

Many program administrators felt very strongly that the practice of measuring free riders is flawed as a whole. “Did we cause the savings? That’s a ridiculous question,” said one program manager. “We don’t believe in free ridership,” said another. “We believe we influence every project we work on. Even if they were already thinking about doing something, we think our involvement caused it to be a better or faster project than it would have otherwise been." On the other hand, evaluators are paid to questions such statements and subject them to a burden of proof, no matter how imprecise.

To be sure, free ridership analysis need not be complicated and resource-intensive. There is evidence that reasonable levels of free ridership can be assessed without using an undue amount of resources.

In Practice

Largely due to the aforementioned common regulatory criteria and the regional planning efforts in the Pacific Northwest, the Energy Trust of Oregon (ETO) notes that their evaluation of free riders is done, "where feasible and reasonable in cost." The best practice here is the ETO's consistent work to educate their regulators and board evaluation committee. Net savings is reported, but the ETO does not ask their contractors to worry about free riders and rewards them based on gross savings. The ETO has worked to convey to stakeholders that while free riders are important, and they will assess and report free riders in program evaluation, that "does not mean that precision is a feasible or useful goal" (Gordon 2012).

In this way, ETO's evaluation of free riders is accepted as an imprecise exercise that can still prove useful for planning and performance assessment purposes, and contractors are given a degree of certainty by not having to worry about which of their customers will be later found to be a free rider. Qualitative assessments, such as whether customers describe ETO's influence as "important" versus "critical" help the ETO determine their value and whether the customer's savings are truly additional to what would have happened anyway.

SPILLOVER

Spillover effects are energy savings that occur when a program indirectly influences energy-saving measures or behaviors in non-program participants. This can occur when a program has influenced vendors, competitors, or suppliers who then influence non-participating companies to make energy efficiency investments. Spillover can also occur inside program participant facilities, where additional energy efficiency projects or practices are undertaken in addition to the ones specifically encouraged by the program. This is called "inside spillover."

One type of spillover effect is that of *free drivers*, who are "those who adopt an energy efficient product or service" due directly to the activities of the program, "but are difficult to identify either because they do not collect an incentive or they do not remember or are not aware of exposure" to the program or program activities (GDS 2007). While some programs use the terms "spillover" and "free driver" interchangeably, spillover is generally viewed as a broader term that can reflect all the additional market impact a program might have, both directly and indirectly. Few programs, if any, appeared to single out free drivers specifically. Most appeared to measure a more general spillover effect, if one was measured at all.

Though 2/3 of states across sectors reported accounting for free riders in their net savings estimates, only 44% reported that there was an adjustment in claimed savings for spillover effect (Kushler et al. 2012). Among industrial programs, 85% of online survey respondents reported evaluation of spillover effects within evaluation activities, though most indicated that spillover effects were only measured occasionally and often not taken into account in net savings.

When spillover was taken into account for net savings, online respondents indicated applied spillover rates of everywhere from .30 to less than .01, with most indicating default spillover rates of less than .05. Several programs indicated that spillover was recently determined as a useful area of pursuit for future evaluations, but that no clear definitions or ratios had yet been determined.

Stakeholder Perspectives

While many respondents indicated that they believe spillover to be equal to or greater than free ridership, few programs are going to great lengths to track it. Much of this is due to the fact that identifying and interviewing or surveying non-participants is much more time-consuming than surveying participants. Short of a census of every industrial firm in a service area, there is no failsafe way to ensure that each major spillover impact is measured. Conducting a statistically significant survey of non-participants is a common way to measure spillover, but such activities are rarely undertaken in evaluation activities. Instead, some programs see their net savings increase by a small amount to account for program spillover.

The general sense among program managers is that significant resources go toward accounting for free riders, while very few are put toward accounting for spillover. The net effect is that substantial free riders are found while little spillover is identified, and net savings are estimated to be much lower than gross savings.

"It's just not something people put substantial resources towards assessing, but there is potential for it to be quite significant. We think it would be useful to put more resources towards trying to determine what that number is," said one policy maker. However, the cost of conducting substantial detailed spillover analysis could be quite high. One program manager echoed many others when he said that his regulator has said to him: "If you think spillover is big, go find it. But we won't look for it," presumably due to a lack of resources and interest. Programs typically reported that they do not have sufficient resources themselves to seek spillover numbers with great precision. It is as much an issue of infeasibility as cost. Evaluators have reported difficulty estimating spillover with even the rough accuracy of free rider estimates because additional actions tend to be poorly documented, are often different from program actions, are not subject to quality control, and are sometimes out of the service territory of the utility paying for the programs.

The benefit of improved spillover analysis is that programs can be more accurately assessed for their impacts overall, and policy makers and system planners can better understand the kind of market transformation occurring within certain sub-sectors or with certain technologies. Such information could be leveraged by program managers to strengthen existing programs while also pointing to emerging opportunities or trends that could help shape future program offerings.

NON-ENERGY BENEFITS

Non-energy benefits are those benefits other than energy savings that accrue to program participants. These can be specific resources, such as water savings, or slightly less tangible items, such as improved worker safety and comfort. In the industrial sector, non-energy benefits are often identified in the form of improved productivity, quality control, or other benefits related directly to the facility's product lines.

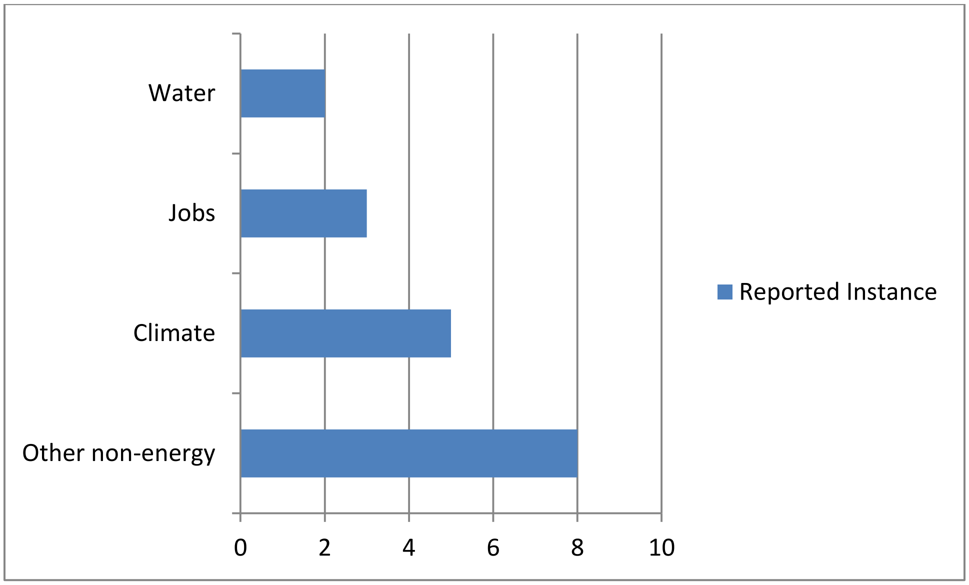
Research consistently suggests that non-energy benefits in the industrial sector are substantial. An analysis of non-energy benefits in Wisconsin found that in calendar year 2010, $8.9 million in non-energy benefits were enjoyed by participants in Focus on Energy business programs, above and beyond the estimated $56 million in annual energy savings for the same year's business customers (Tetra Tech 2011). Often non-energy benefits are far greater than any energy benefits provided by an energy efficiency measure (see Lung et al. 2005, Worrell et al. 2003).

In the survey of energy efficiency programs across all sectors, 30% of states indicated that they quantify non-energy benefits to society and/or customers in their evaluations and cost tests, though quantified benefits to customers were typically limited to water and other fuel savings (Kushler et al. 2012). Most states did not measure any non-energy benefits. As Kushler et al. note, these types of non-energy benefits to customers are typically not measured or estimated, while all customer costs *are* often included in program cost tests (Kushler et al. 2012).

Unfortunately, few programs themselves reported conducting substantial analyses of non-energy benefits. States that use the total resource cost (TRC) cost tests—that is, the majority of states surveyed by Kushler et al.—do not include non-energy benefits in their cost test calculations. Industrial projects can yield high levels of non-energy benefits, which, if included in a cost test, could increase rates of project acceptance and yield greater energy efficiency within the sector (Kushler et al. 2012).

In the industrial online survey, about half of the respondents indicated that specific non-energy benefits were measured for their programs. Figure 5 highlights the areas in which non-energy benefits were reported to be measured for online survey respondents. While it is useful to see the number of programs that do measure these benefits, it is important to note that survey respondents generally indicated that there is little effort put into measuring these benefits, and in most cases there is no established methodology for measuring them. These additional benefits should be considered to be only minimally measured.

Figure 5. Types of Non-Energy Benefits Included in Evaluation Activities



Note: Respondents were able to choose more than one answer
Source: Online survey

Stakeholder Perspectives

Not only do non-energy benefits appear to be largely left out of cost tests, they often serve to reduce the savings for which a program can take credit. Some program managers noted that evaluation efforts to ascertain non-energy benefits were used to determine program influence and measure for free riders. If customers report substantial non-energy benefits, they may be asked whether those benefits caused the investment to be made. They may indicate that the non-energy benefits (such as increased productivity) where the most important reason the investment was made, resulting in the customer being labeled as a free rider for purposes of the energy efficiency program. As a result, some program managers feel that discussing non-energy benefits as marketing tools can backfire on them during the evaluation process.

Non-energy benefits remain a poorly understood aspect of industrial energy efficiency measures, in part because they are rarely studied. Only five programs reported measuring for climate-related impacts, and only three reported measuring for jobs impacts. Both of these types of non-energy benefits could be of increasing importance in the coming years, and it would behoove policy makers and regulators to better understand the ancillary benefits of energy efficiency programing. As one program manager said, "I believe non-energy benefits are an order of magnitude bigger than energy savings." Energy efficiency programs are well positioned to quickly enhance everyone's understanding of the additional benefits of energy efficiency.

Despite the fact that non-energy benefits might be a driving factor encouraging an industrial firm to make an energy-efficiency investment, their absence from TRC tests ensures that utilities will continue to discount them. This is highly problematic for utilities interested in maximizing industrial energy efficiency potential, because industrial projects that could be highly cost-effective due to their non-energy benefits will not be identified and encouraged.

Regulators and evaluators may be hard-pressed to encourage substantial evaluation of non-energy benefits, because they are tasked specifically with evaluating *energy* programs. Policymakers could help to explicitly encourage the evaluation of non-energy benefits and recommend that evaluation resources be dedicated to such assessments.

Additionally, the emphasis on high levels of precision in all aspects of evaluation may be driving up the cost of evaluation and making it even harder to justify additional resources to measure other impacts such as non-energy benefits. Encouraging industry-wide or utility-wide acceptance of certain reasonable levels of precision for all effects might free up some evaluation resources to put toward the measurement of previously under-measured effects.

The Future of Industrial Program Evaluation

The manner in which industrial energy efficiency programs have been evaluated in the past will need to change to address and respond to current and future changes in the way energy resources are identified, planned for, paid for, and assessed. The following section will discuss some of the changes happening now and on the horizon in industrial energy efficiency that evaluators will be called on to address.

ATTRIBUTION AND CLIMATE CHANGE

Energy efficiency programs have largely existed to meet specific energy efficiency goals and to provide reliable and low-cost energy supplies. The emissions benefits and other ancillary benefits conveyed by energy efficiency were not generally the rationale behind energy efficiency investments.

However, in recent years, things have changed. With an increasing emphasis on climate change mitigation and greenhouse gas reduction, energy efficiency will play an ever-increasing role in emissions reduction strategies. New federal rules impacting existing and new electrical generation resources are explicitly looking to energy efficiency resources to meet emission reduction goals. Absent federal leadership on carbon dioxide emissions reduction schemes, some states are developing their own long-term carbon reduction strategies and plans. Again, energy efficiency resources are seen as cost-effective, clean ways to meet future energy needs while satisfying climate-related goals.

As these emission reduction programs enter and influence the marketplace, it will become harder and harder to determine which programs and policies influenced which energy efficiency investments, and to what degree. The study of determining which programs have which influence is called attribution. Program managers and evaluators both agreed generally that attribution is becoming much harder to solve for.

Recent interactions of utility and PBF programs with programs funded by the federal *American Recovery and Reinvestment Act of 2009 (ARRA)* highlight the challenges ahead. While some states gave full credit to utilities and PBF organizations for measures implemented with help or support from ARRA funds, other states developed credit schemes that awarded portions of savings to certain programs. As one evaluator put it, "if you have a big project that was influenced by four distinct programs, and you divide the savings four ways, everyone gets a net-to-gross ratio of .25 on those savings. That looks bad, and that is not going to pass internal cost tests in some cases, though it would have easily if the project savings were just attributable to one program."

Projects may be encouraged by a multitude of programs that, when combined together, cause the investment. However, one program individually may not have such an impact. The synergy of the programs together may encourage the investment. Fighting over which program gets to claim which kWh misses that point that sometimes projects only happen only because of the influence of multiple programs. This conflict is exacerbated by shareholder incentives or lost-revenue adjustments, which accrue only when specific savings can be proven or claimed by a program.

As currently structured in states which do not acknowledge that multiple parties can play critical roles in achieving the same savings, the fight over energy savings is effectively a zero-sum game. It can hamper creativity among program managers and designers if they become worried that their savings from more nebulous or behavior-focused programs could possibly be attributed to other programs. It could also reduce interest in partnerships between complimentary agencies and entities if programs begin to feel that they must guard their customers and their "claim-able" measures from outside influences.

ENERGY MANAGEMENT ACTIVITIES

One of the major changes in industrial energy efficiency programming that evaluation must respond to is the rising importance and development of energy management activities and programs. These programs encourage strategic energy management, continuous improvement, and overall energy performance. They include the use of tracking tools to enable real-time tracking of energy use; substantial behavioral changes; intra-industry benchmarking and energy goal-setting; and significant outreach and interaction with vendors and other third parties able to provide specific services and tools.

Energy management programs are capable of totally changing the paradigm of how industrial energy efficiency programs are evaluated. This is due to the fact that energy management programs, once put in place, consistently identify and assess energy efficiency opportunities, resulting in the continued optimization of a facility's energy performance. Whether energy savings occurs as a result of a certain piece of equipment or a change in behavior or process ceases to matter to the program. As one energy management expert described, "We want to get to the point where facilities have energy management systems that are functioning, that conform with international standards, and are producing savings verifiable by a third party. If the savings is verified, it doesn't matter where they came from. It just matters that you can measure it against the baseline. You know exactly what you're saving."

This holistic approach to energy programming and measuring will be hard to marry with the typical existing approaches to industrial evaluation. Metrics such as energy use per unit of production might need to become more common in evaluation approaches in order to better reflect the impact of energy management activities. The upside is that these energy management activities are producing, and will continue to produce, levels of data on measure performance and industry practices that have not previously been available. With every new facility that begins to more diligently track their energy savings and meter specific aspects of their operations, deeper understandings of how systems within facilities use energy will be available. The challenge will be to help companies feel comfortable with sharing this data.

Some energy analysis firms offer approaches to meshing metered data, billing data, and other data to develop plant-specific models that can give program administrators, evaluators, and regulators a much clearer sense of how a facility has improved its energy performance over time. Energy management programs will continue to encourage this kind of tracking of energy performance against a baseline and the integration of other data and improved modeling of energy performance. The challenge for evaluators, regulators, and policymakers, is to understand and feel comfortable with the data and the models being put into use today. They will need to concurrently become familiar and comfortable with the idea that changes in behavior and operations can indeed yield substantial, creditable, and persistent savings.

At some point, utilities all over the country will likely have energy management offerings. So too will take-up of various energy or emissions goal-setting and external certification of energy performance rise. There will be large differences between truly high-performing entities and the mere presence of corporate energy-saving goals or certification within an energy standard. Evaluators will be called upon to make those distinctions.

They will also be called on to evaluate the true persistence of energy-saving behaviors and operational changes. Once customers are continually tracking energy use and making adjustments to their operations, the initial influencers will wish to claim additional savings as their own. In addition to questions of attribution, regulators and policy makers will want to know whether these behaviors are being maintained. With high-performance energy management activities new to many facilities and sub-sectors, the evaluation community needs time to have a good sense of persistence and general maintenance of the savings (Eckman 2012). As years progress beyond when the initial baseline was developed, evaluators and regulators may have a difficult time accepting previously derived baseline data as relevant for new evaluations.

Regulators especially will be called on to approve and support the development of energy management programs. These programs are a substantial change from legacy energy efficiency programming, and, as one evaluation expert noted, "If operations, and cleaning them up to be higher energy performing, are seen as a somehow less worthy activity by regulators, it will be difficult to fully recognize the opportunity."

Self-Direct and Opt-Out Programs

As noted in previous ACEEE research, self-direct programs, which allow industrial firms to self-direct a portion of their PBF funds, are increasingly being developed in states and provinces in the U.S. and Canada. So too has the number of opt-out programs, which allow industrial firms to opt-out of paying all of their PBF funds, risen in recent years (Chittum 2011). Opt-out programs require no evaluation whatsoever of the savings claimed by participating facilities. They are a poor use of PBF funds and should not be encouraged as a policy option. Self-direct programs typically require some oversight of claimed savings, but only a small number of existing self-direct programs actually require the same level of evaluation, measurement and verification as other utility and PBF organization efficiency program offerings (Chittum 2011).

If industrial customers continue to ask for and be granted opt-out or self-direct options, as they have in the recent past, the amount of claimed energy savings not subject to appropriate oversight will likely increase. With industrial energy efficiency such a cost-effective resource, regulators and policy makers cannot afford to let additional claims of energy savings in the industrial sector be made absent sufficient evaluation, measurement and verification. The handful of self-direct programs that *do* require substantial oversight and evaluation of claimed savings are evidence that self-direct programs can be developed in a manner that ensures PBF funds are used fairly and prudently to acquire cost-effective savings that benefit the entire energy system (Chittum 2011).

Discussion

The Role of Regulators

Much of this report has focused on the opinions and activities of evaluators and program managers. But an overarching theme of interviews and surveys conducted for this report was the strong influence regulatory commissions have on the shape of evaluation.

Regulators have a unique ability to shape energy efficiency programming and respond to new trends and challenges. They approve program designs and help define the evaluation requirements. They can recognize when the overall policy aims of energy efficiency programming are not well reflected in the current approach to evaluation. They can address the efficacy and necessity of various degrees of precision.

For instance, as mentioned earlier, some programs may be running with high rates of free riders but substantially larger spillover impacts. Regulators can help encourage additional measure of spillover in instances where spillover impacts are not adequately measured. Otherwise such impacts will remain unknown, and utilities might be encouraged to discontinue programs that are effecting desirable changes in energy use.

Regulators could encourage better and more meaningful analysis of impacts in areas where the barrier between customer and non-customer is fading, as in education, training, and vendor outreach efforts. They could also assess the value of certain evaluation data points, and consider whether more qualitative analyses could be more useful to program designers looking to ascertain nuances of their program. One program manager suggested, "I would favor some information gathering that gets at a better understanding of the decision-making process, such as asking questions in the survey like 'What information played an important role in your decision making process?' Then present the customer with some options." Asking them if it was the program's cash flow analysis, technical analysis, customer service, etc., could help illuminate issues of free ridership while also providing highly useful feedback to program managers and designers about their customers' decision-making activity.

Similarly, program managers note that reducing the emphasis on net savings could still yield useful data to measure program performance while also freeing up resources to measure other important but understudied metrics. As noted by one program manager, "In some states, instead of the burden of proof lying on programs to show savings, regulators presume the programs are getting the savings unless presented with evidence to the contrary." For purposes of system planning, ensuring accurate adjusted gross savings and then stating that a certain percentage of participants are *likely* free riders is one way to still address free ridership while not removing a specific, contested amount of actual kWh savings from a program's savings claims. Program performance incentives could or could not be tied to net savings, but they could also be tied to gross savings while regulators keep an eye on free ridership.

Regulators will play a critical role in shaping future industrial energy efficiency programs. The manner in which these programs are evaluated will in turn shape the way industrial energy efficiency is viewed as a resource by policy makers around the country and the world. With increased customized approaches to acquiring energy efficiency, evaluation activities and regulator will need to keep pace with these changes and recognize the uniqueness of these future efficiency opportunities.

CONCLUSIONS AND NEXT STEPS

Energy efficiency, and industrial energy efficiency in particular, has consistently proven to be one of the most cost-effective energy resources available to states to meet future energy demands. As a result, energy efficiency program spending has consistently risen in recent years. Industrial energy efficiency is a prioritized energy resource, and programs run by utilities and PBF organizations continue to be called on more than any other entities to acquire it (Chittum and Nowak 2012, Friedrich et al. 2009, Lazard 2011).

Evaluation activities will increasingly be called upon to determine the impact of these programs and to paint a clearer picture of how industrial energy efficiency programs interact with other energy and non-energy programs with similar or complimentary goals. Evaluation is a very important aspect of energy efficiency program administration, and helps ensure that public benefits are being acquired in a cost-effective manner. Evaluation can provide a more unbiased assessment of program impacts than a program administrator might themselves, and can help identify areas of the program that need more attention. However, evaluation approaches of years past may not be appropriate or adequate to fully identify and highlight the impacts of newer energy efficiency programming.

ACEEE recognizes the unique circumstances and situations in which utilities and PBF programs operate from state to state and service territory to service territory. While blanket recommendations for how to evaluate industrial energy efficiency programs are not appropriate, suggestions for how to improve evaluation can be made. ACEEE suggests that regulators, program administrators, and evaluators:

- Consider measuring net-to-gross less frequently than annually if the evaluation resources could be better used to measure other effects, and measure only to one significant figure;
- Assess the extent to which the timing of various evaluation activities is meeting the needs of all major stakeholders, including industrial customers;

- Determine whether the measurement of free riders is adequately balanced by the measurement of spillover impact, so as to better assess the full influence of a program;
- Increase efforts to measure non-energy benefits, so as to more accurately reflect the impact of industrial energy efficiency programs and perhaps yield data that programs could use to better market the programs to customers; and
- Remain open to the fact that evaluation approaches may need to change as utilities and PBF organizations begin to offer more energy management programs and other next generation programs that mesh resource acquisition with market transformation.

In working toward the above suggestions, some larger questions could help guide the discussion among all interested stakeholders. These include:

1. What are our energy efficiency policy goals, and is evaluation helping us achieve them?
2. If evaluation is producing results that are discounted or dismissed by program managers, how much value is it really providing? How could it be improved?
3. Are the metrics sought by regulators and evaluators telling the best and most complete story about the impact of industrial energy efficiency programs?
4. What level of precision in evaluation activities is warranted? Are requirements for high levels of precision in measuring something that cannot be perfectly measured creating false precision?
5. If a high degree of precision is not always needed, would "good enough" assessments of program impacts suffice? What would those look like?
6. Are evaluation reports easy to understand by stakeholders? How can we produce reports that are useful and accessible to program and system planners?

These conversations need to involve program managers, evaluators, regulators, policy makers, third-party engineers, and the industrial customers who will be making energy efficiency investments. Recognizing that fighting over kWh and dollars is unproductive, these parties should work together to find a compromise that results in reasonable certainty of program performance towards mutually agreed-upon goals.

Evaluation should give program implementers, designers and policy-makers useful information about the impact of their program and possible program changes in the future. It should allow program managers and administrators to feel that their programs are fairly evaluated and that the most useful data points are being collected in an accurate manner.

Evaluation data is useful to parties beyond regulators. Certain jurisdictions reported that they expect to increasingly rely on their industrial sectors to meet future savings goals than in years past. Industrial efficiency supply curves used to forecast future industrial energy efficiency resources are generally more difficult to estimate due to the confidential nature of much industrial data. These supply curves must rely on many assumptions due to a lack of widely available data on industrial efficiency investments. Data from industrial energy efficiency programs, including energy management programs and behavioral-focused approaches, will be increasingly needed by system planners to improve these industrial efficiency supply curves. Meaningful evaluation could help bring to light a lot of this data.

The value of the various metrics collected by evaluators should be routinely assessed. Collecting easily available data just for the sake of collecting data confuses regulators and policy-makers and leaves other metrics unstudied. As one report on measuring net savings in the Northeast noted, "There is a danger that the effectiveness of programs could be reduced if the energy efficiency community focuses program efforts only on what is most provable, not what is most effective" (NMR 2010). Evaluation has a strong influence on the way programs are run, and if evaluation discounts or generally ignores the benefits of new and promising programs, those programs may not continue.

What industrial energy efficiency programs and influences will look like in five, ten, or twenty years is hard to know. But the manner in which evaluation is conducted and findings are utilized will either encourage or discourage program creativity, new partnerships, and overall performance. The questions of influence, of attribution, and of value are becoming increasingly nuanced, especially in the industrial sector. Evaluation will need to adapt to these changes to remain relevant, useful, and fair. While some programs have already begun to alter and update their evaluation approaches, most programs appear to be unhappy with the manner in which their programs are evaluated.

Ultimately, everyone in the industrial energy efficiency world has a common goal. All parties need to feel that evaluation results are credible and useful. If stakeholders are fighting over evaluation results, the important story—that industrial energy efficiency is a highly cost-effective and beneficial energy resource—will be lost.

References

[ACEEE] American Council for an Energy-Efficient Economy. 2011. *State Energy Efficiency Resource Standard (EERS) Activity*. http://www.aceee.org/files/pdf/policy-brief/State%20EERS%20Summary%20October%202011.pdf. October. Washington, DC: American Council for an Energy-Efficient Economy.

Blumstein, Carl. 2009. *Program Evaluation and Incentives for Administrators of Energy-Efficiency Programs: Can Evaluation Solve the Principal/Agent Problem?* http://www.ucei.berkeley.edu/PDF/csemwp184.pdf. Berkeley, CA: University of California Energy Institute, Center for the Study of Energy Markets.

[CEE] Consortium for Energy Efficiency. 2012. *State of the Efficiency Program Industry: Budgets, Expenditures, and Impacts 2011*. Boston, MA: Consortium for Energy Efficiency.

Chittum, Anna. 2011. *Follow the Leaders: Improving Large Customer Self-Direct Programs*. Washington, DC: American Council for an Energy-Efficient Economy.

Chittum, Anna and Seth Nowak. 2012. *Money Well Spent: 2010 Industrial Energy Efficiency Program Spending*. http://aceee.org/research-report/ie121. Washington, DC: American Council for an Energy-Efficient Economy.

Chittum, Anna, Neal Elliott, and Nate Kaufman. 2009. *Trends in Industrial Energy Efficiency Programs: Today's Leaders and Directions for the Future*. http://aceee.org/research-report/ie091. Washington, DC: American Council for an Energy-Efficient Economy.

Clarke, Leon, Fred Coito and Frank Powell. 1996. "Impact Evaluation of Pacific Gas & Electric's Industrial Process, Refrigeration, and Miscellaneous Measure Program." In *Proceedings of the 1996 ACEEE Summer Study on Energy Efficiency in Buildings.* http://aceee.org/proceedings-paper/ss96/panel06/paper05. Washington, DC: American Council for an Energy-Efficient Economy.

Coito, Fred (KEMA). 2012. Personal communication. April 3.

Crossman, Kim (Energy Trust of Oregon). 2012. Personal communication. June 13.

Dodenhoff, James (Southern California Edison). 2012. Personal communication. May 10.

Eckman, Tom (Northwest Power and Conservation Council). 2012. Personal communication. June 14.

———. 2011. *Integrating Energy Efficiency into Resource Acquisition in the PNW: The Role of EM&V*. Northwest Power and Conservation Council.

[ECW] Energy Center of Wisconsin. 2012. "Overview of Cost-Effectiveness Tests." *Energy Efficiency Guidebook*. http://www.ecw.org/publicpowerguidebook/content.php?chapterid=9&sectionid=9. Accessed June 5. Madison, WI: Energy Center of Wisconsin.

[ETO] Energy Trust of Oregon. 2011. *True Up 2011: Tracking Estimate Corrections and True Up of 2002-2010 Savings and Generation.* http://energytrust.org/library/reports/2011_TrueUpSummaryPaper_Final.pdf. Energy Trust of Oregon.

[FEMP] Federal Energy Management Program. 2008. *M&V Guidelines: Measurement & Verification for Federal Energy Projects, Version 3.0*. http://www1.eere.energy.gov/femp/pdfs/mv_guidelines.pdf. Washington, DC: Department of Energy, Federal Energy Management Program.

Friedman, Rafael (Pacific Gas and Electric). 2012. Personal communication. February 23.

Friedrich, Katherine, Maggie Eldridge, Dan York, Patti Witte and Marty Kushler. 2009. *Saving Energy Cost-Effectively: A National Review of the Cost of Energy Saved Through Utility-Sector Energy Efficiency Programs*. Report U092. Washington, DC: American Council for an Energy-Efficient Economy.

[GDS] GDS Associates, Inc. 2007. *Vermont Electric Energy Efficiency Potential Study*. Prepared for the Vermont Department of Public Service. GDS Associates, Inc.

Glanton, Cheryl (New York State Energy Research and Development Authority). 2012. Personal communication. March 30.

Goetze, Karl (Efficiency Vermont). 2012. Personal communication. April 4.

Gordon, Fred (Energy Trust of Oregon). 2012. Personal communication. April 6.

Kushler, Marty (American Council for an Energy-Efficient Economy). 2011. Personal communication. December 12.

Kushler, Martin, Seth Nowak and Patti Witte. 2012. *A National Survey of State Policies and Practices for the Evolution of Ratepayer-Funded Energy Efficiency Programs*. http://aceee.org/research-report/u122. Washington, DC: American Council for an Energy-Efficient Economy.

Lazard. 2011. "Levelized Cost of Energy Analysis -Version 5.0." http://energy-ecology.blogspot.com/2012/03/levelized-costs-of-electricity.html. New York, NY: Lazard, Ltd.

Lung, Robert Bruce, Aimee McKane, Robert Leach, and Donald Marsh. 2005. "Ancillary Savings and Production Benefits in the Evaluation of Industrial Energy Efficiency Measures." In *Proceedings of the 2005 Summer Study on Energy Efficiency in Industry.* Washington, DC: American Council for an Energy-Efficient Economy.

Maxwell, Jonathan B., Nikhil Gandhi, and Fred Coito. 2011. "Establishing Baselines for Industrial Energy Efficiency Projects." In *Proceedings of the 2011 Summer Study on Energy Efficiency in Industry.* Washington, DC: American Council for an Energy-Efficient Economy.

McKane, Aimee (Lawrence Berkeley National Laboratory). 2012. Personal communication. April 17.

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

[NAPEE] National Action Plan for Energy Efficiency. 2007. *Model Energy Efficiency Program Impact Evaluation Guide*. Prepared by Schiller Consulting, Inc. http://www.epa.gov/eeactionplan. Washington, DC: U.S. Environmental Protection Agency.

[NMR] NMR Group, Inc. 2010. *Net Savings Scoping Paper*. http://neep.org/uploads/EMV%20Forum/EMV%20Products/FINAL%20Net%20Savings%20Scoping%20Paper%2011-13-10.pdf. Northeast Energy Efficiency Partnerships.

[NWPCC] Northwest Power and Conservation Council. 2010a. *Sixth Northwest Conservation and Electric Power Plan*. http://www.nwcouncil.org/energy/powerplan/6/default.htm. Northwest Power and Conservation Council.

–––. 2010b. *Briefing Book*. http://www.nwcouncil.org/library/2010/2010-13.htm. Northwest Power and Conservation Council.

Romero, Kenny (Xcel Energy). 2012. Personal communication. June 26.

Tetra Tech. 2011. *Focus on Energy Evaluation Annual Report (2010)*. http://www.focusonenergy.com/files/Document_Management_System/Evaluation/annualreport2010_evaluationreport.pdf. Tetra Tech.

Wallner, John (Northwest Energy Efficiency Alliance). 2011. Personal communication. November 2.

Worrell, Ernst, John A. Laitner, Michael Ruth & Hodayah Finman. 2003. "Productivity Benefits of Industrial Energy Efficiency Measures." *Energy* 11 (28)1081-1098.

Appendix I: Telephone Interview Framework

These questions guided the telephone interviews conducted with program administrators, regulators, and evaluators.

1. Who are the principal parties involved in evaluation? What is everyone's role and responsibility?
2. Is there one clear protocol for the whole state, or does each utility do it differently?
3. Are there differing levels of evaluation based on types/sizes of projects?
4. What is the time frame: what happens first, when is baseline developed, when do evaluators first come in?
5. How are specific sites chosen for additional evaluation, surveys, or interviews? If large projects, how are they defined?
6. Do you develop a measure of the facility's baseline? How do you do that?
    a. How does your methodology for measuring a baseline reflect or respond to changes in production lines, shift changes, changes pre/post installation?
7. What is the relationship between impact evaluation and program M&V?
8. What kind of specific data are you collecting from each site?
    a. Who gets to get into the site? What do they get access to?
    b. What kind of specific inputs are you collecting from each site?
    c. What additional data do you wish was collected?
    d. Do you rely on internally produced engineering analyses? Are these analyses evaluated by anyone else?
9. Are you developing sample realization rates and applying that to all projects?
10. In general, how do you measure gross energy savings at each site?
    a. When do you use which approaches?
11. In which instances and for what situations do you use savings estimates?
12. Do you determine net or gross savings?
    a. Why?
    b. Have you come up with numbers you believe are accurate?
    c. (If applicable) What additional data would help feel more confident in net savings numbers?
    d. What kinds of interviews/surveys are conducted? Who does them?
    e. What do you (or others) use net or gross numbers for?
13. If you determine net savings:
    a. How do you define it?
    b. Is there a set net-to-gross ratio applied?
14. How do you determine free ridership?
    a. How accurate do you think those numbers are?
    b. Is a default value used?
    c. Do you think the methodology for measuring free ridership aligns with the actual internal decision-making activities of industrial firms?
15. What kinds of evaluation activities have yielded credible results that you've taken action on? Has that action improved what you do?

16. What have you taught regulators about your programs? What have they become comfortable with?
17. Do you evaluate any behavior or O&M-focused activities? How?
18. Do you measure non-energy benefits? How? Which benefits?
19. Do you measure non-participant spillover effect of any programs?
20. Are you concerned with attribution to any other energy efficiency or climate change programs?
21. Is the cost-effectiveness of evaluation itself ever measured or assessed?
22. Are you evaluating industrial self-direct programs as well? What does that entail?
23. What cost tests are you using?
24. Do you think your evaluation process is fair? What would you change?
25. Do you think the goals of the evaluation process are well aligned with those of the program administrators?