# Kicking the Can: How First-Year Impact Evaluation Transfers Cost and Uncertainty

*Marc Collins and Joe Loper, Itron, Oakland, CA*

## ABSTRACT

Evaluation is now standard practice for all utility energy efficiency programs, whether small or large, new or mature, boiler-plate or innovative. A discrete set of policy interests are driving increased interest in evaluation, including: assessment of program cost effectiveness (almost always including a total resource cost test), compliance with state energy efficiency resource standards (EERS), and bidding of savings for carbon offsets.

Unfortunately, most utility evaluations are not necessarily collecting all or the best data needed to address these policy interests, which are long-term in nature. To understand whether multi-year and sometimes multi-decade policy goals are being met, credible and reliable multi-year evaluation studies that account for variability of energy efficiency measure and program persistence, as well as estimation of impacts that would otherwise not occur due to market forces or other factors already included in reference load forecasts are required. Instead, most evaluations are limited to "first-year savings."

Limiting evaluation scope to first-year savings usually avoids questions of measure persistence; current standard practice is to use estimated useful life estimates that were, for the most part, developed more than a decade ago. Evaluation of first-year savings also avoids questions of multi-year baselines; current typical evaluation practice is to assume first-year baselines throughout measures' assumed lives. In addition, it usually avoids collection of data related to incremental costs.

Evaluations of first-year savings can appear less costly and avoid the need for up-front discussion of key uncertainties associated with the use of evaluation results. But, more often, simply transfer the costs and uncertainty to the post-evaluation analyses. Moreover, through delay, it adds to the costs of obtaining the data needed to estimate life-cycle savings and costs. In the process, the uncertainties associated with cost effectiveness analysis, compliance with EERS, and other issues are often obscured, since the focus is on high relative precision achieved for key first-year savings parameters with limited, if any, discussion of uncertainties of other parameters needed to complete cost effectiveness or valuation analyses.

This paper briefly examines evaluation drivers and objectives in the current context of the increasing emphasis on energy efficiency spending; reviews state and utility evaluation plans and assesses the extent to which their focus is on first-year impact evaluations; discusses why jurisdictions are relying on first-year savings rather than collecting other data and performing lifecycle analysis; and, identifies alternative approaches to evaluation studies that would better integrate a broader range of data and analytical requirements needed to address longer-term policy questions, taking into account related data challenges.

## Evaluation Drivers and Objectives

Credible estimation of the impacts of energy efficiency programs is a critical element of state, regional and national efforts to make energy efficiency programs a reliable, economical and significant part of the energy resource mix for decades to come. Evaluation is now standard practice for utility energy efficiency programs.

- 40 of the 45 states (including the District of Columbia) that have formally approved utility energy efficiency programs have legislation or regulatory orders requiring utilities to evaluate their programs (Kushler et al. 2012, p. 9)

- 30 states use the evaluated savings estimates for some form of utility compensation determination, including shareholder incentives, program cost recovery, and lost revenue recovery (Kushler et al. 2012, p. 10)
- 35 states have legal requirements for the use of benefit-cost tests for ratepayer-funded, energy-efficiency programs (Kushler et al. 2012, p. 11) and 5 more states at least consider cost effectiveness in program and/or portfolio design
- 29 states use the Total Resource Cost (TRC) Test as their primary test, while six use a Societal Cost Test (SCT), five use a Utility Cost Test (UCT),[1] and one uses a Ratepayer Impact Measure (RIM) test (Kushler et al. 2012, p. 13)
- 24 states have established energy efficiency savings targets, commonly referred to as energy efficiency resource standards (EERS) or energy efficiency performance standards, that utilities or non-utility program administrators (PAs) must meet through customer energy efficiency programs.[2]

State EERS may prescribe specific energy savings targets or, in some states, the acquisition of "all cost effective energy efficiency." Energy efficiency is also an eligible (or mandatory) resource in some Renewable Portfolio Standards (RPS). Where energy efficiency is "allowed into the tent" there is a secondary interest in impact and cost-effectiveness analysis vis-à-vis supply (usually long-term focused) versus demand (usually short-term focused) resources. Renewable energy advocates are sometimes concerned that including energy efficiency program savings in the cap could reduce the effective stringency of the renewable energy requirement and, consequently, the environmental and other benefits, if the energy efficiency program savings are not (perceived to be) real and persistent.

Some utilities are using energy efficiency program evaluations to support bids to provide capacity resources into regional power systems, such as the PJM and ISO-New England. This suggests that energy efficiency program savings are, at least to some degree, being considered system resources by power system planners.

Planners use "top-down" econometric forecasts and jurisdiction-wide assumptions for integrated resource planning that could potentially be informed by "bottom-up" results from program evaluation studies. Analyzing whether the two "meet in the middle" is a key step for planners where DSM is being taken seriously. The key to such a paradigm change is credible and reliable multi-year evaluation studies that account for variability of energy efficiency measure and program persistence, as well as estimation of impacts that would otherwise not occur due to market forces or other factors already included in reference load forecasts..

## First-Year Savings Evaluation Issues

At least initially, for a variety of reasons, most utility energy efficiency program evaluations focus on first-year, as opposed to multi-year, energy and demand savings. However, if evaluations are not designed to address the multiple drivers and objectives outlined in the section above, the results could be incomplete, irrelevant, inaccurate and/or misleading. Three key issues are often down-played or ignored in first-year savings evaluations: persistence of measure savings, the potential for changing annual savings estimates due to changes in baseline usage over time, and the incremental costs related to installing measures and early replacement of equipment.

[1] Sometimes referred to as a Program Administrator Cost Test.
[2] ACEEE website, http://aceee.org/topics/eers (accessed April, 2013)

## Measure Persistence

First-year savings evaluation typically ignores the persistence complication. Many evaluation processes undertaken today use estimated useful life (EUL) estimates that were, for the most part, developed a decade or more ago. Many recent persistence studies are either narrowly focused or compilations of secondary data (Jayaweera et al. 2013, Ch. 13), but some promising exceptions exist along with a recent wave of studies of behavioral programs.

To reflect reality, savings from previous years' activity should ideally be adjusted, as required, each year to reflect (differences in) measure failure, diminished relative performance, and decommissioning. Especially for short-lived measures, assumptions regarding the EUL of measures could have a large impact on savings in a given year from cumulative program activity. For early replacement retrofit programs, a credible estimate of the remaining useful life (RUL) is also critical. Although the original pre-retrofit equipment would also degrade if left in place, the degradation rates may not be the same.

Measure life assumptions can have material effects on cost effectiveness as well. A large portion of most portfolio savings still come from CFLs, which are short-lived measures. If program bulbs are in sockets for four years, and assumed to be in sockets for six years, the actual benefit-cost ratio will be roughly one third lower than estimated. While that may not result in lighting programs failing a cost effectiveness test, it could deplete the ability of the lighting programs to "carry" other programs that are not standing on their own (in jurisdictions that use sector- or portfolio- as opposed to measure- or program-level cost effectiveness for program and portfolio planning). Especially in the residential sector, non-lighting programs are often reliant on lighting programs to be cost effective at the sector level. Likewise, if the CFLs (or LEDs) are lasting longer than assumed, lighting programs would contribute even more benefits to sector-level cost effectiveness.

Cost trade-offs will necessarily limit the study of persistence for a broad range of measures by any single jurisdiction or utility. However, regional and/or national persistence studies, conducted by the national laboratories or regional groups (e.g., the Northeast Energy Efficiency Partnerships) could help achieve necessary economies of scale for these studies and help to furnish sufficient longitudinal data that is so often a major barrier to persistence studies.

## Multi-Year Baselines

More often than not, evaluation of first-year savings avoids questions of estimating or collecting data to support multi-year baselines.

Baseline conditions—"what would have happened to the equipment in the absence of the program"—are typically defined in one of three ways. An early replacement baseline compares program measure energy consumption and costs to the equipment that was already in place (*in situ*). A standard practice baseline assumes the replaced equipment was at the end of its life or that the measure is part of a new installation or new construction. In the absence of a market study, federal or state equipment (performance) standards and codes are commonly used as a proxy for "standard practice" efficiency levels.

The third baseline definition is called a dual baseline. Dual baselines employ both an early replacement (*in situ*) baseline for the RUL of the original equipment and then apply a standard practice baseline for the remainder of the new measure's EUL. The dual baseline attempts to reflect the normal turn-over of the piece of equipment had it been left to wear-out. The program-induced early replacement of that equipment is given full credit for the period that was truly "early replacement," but reduced savings for the period where replacement would have occurred without the program. Note that it is not assumed that early replacement occurs because of the program—that is an attribution question. End users also sometimes replace functioning equipment early in the absence of programs for a wide variety of reasons (e.g., regulations, changes in market driven production requirements, safety considerations, labor costs, aesthetics, energy costs, etc.).

Early replacement baselines are frequently (and sometimes erroneously) assumed for a variety of measures (commercial lighting is a notable example). In a comprehensive effort, Skumatz (2011) could only find a few primary RUL studies. Assigning accurate values to EUL and RUL is not easy. A first-year savings evaluation model allows this significant savings factor to be pushed off to the indefinite future.

Dual baselines could significantly reduce savings estimates for measures where early replacement (i.e., that the *in situ* fixture would have lasted forever) was previously assumed. On the other hand, using a dual baseline instead of a standard practice (or code) baseline may increase savings in some situations (e.g., if evidence strongly indicates early replacement of equipment that would otherwise remain functioning and *in situ* for some period of time). In addition, even in cases where dual baselines may decrease savings, their proper use may actually increase cost effectiveness due to the relationship among full and incremental costs and the timing of projected equipment replacement.

### Participant Incremental Measure Cost

Since it is not needed to quantify the energy savings, collection of data related to incremental measure cost is typically outside the scope of first-year savings evaluations despite the fact that these are a critical input assumption to the California Standard Practice Manual (SPM)-derived cost-effectiveness tests that are used in most jurisdictions. The quality and accuracy of *ex ante* measure cost estimates therefore have a clear and direct impact on the quality and robustness of the corresponding portfolio-level decisions made by regulators and program planners (Ting et al. 2013).

In fact, only a small minority of Technical Resource Manuals (TRMs) published to date include estimates of incremental measure costs (Ting et al. 2013). A recent comprehensive study of TRMs (Jayaweera et al. 2011) delves into phenomenal detail about methodological differences, but also polled potential users about gaps and needs. Despite its importance for cost-effectiveness calculation, the incremental measure cost issue is not even addressed or mentioned in that study—likely not due to an oversight, but because evaluators rarely have any scope or funding for measure cost estimation. Evaluation managers and efficiency policy makers are generally fixated on savings calculations. This is reflected more broadly in the energy efficiency program industry over its entire history in the huge imbalance in funding levels for estimation of savings and estimation of measure costs. One analyst who led several of the first comprehensive measure cost studies in the 1990s, informally estimates the spending gap for savings as compared to measure cost estimation at on the order of one hundred fold (CPUC 2010 and Rufo 2013). There has been a recent increase in funding for measure cost studies in a few jurisdictions and regions—including California, Massachusetts and the Northeast (NEEP)—but measure cost estimation is still the exception rather than the rule in the evaluation community.

## Analytical and Data Gaps

Narrowing evaluation to first-year savings often fails to adequately inform the various purposes for which the evaluations are intended in the first place. Some key examples where this is evident are outlined below.

### Utility Compensation

Many utilities are required to submit a demand-side management (DSM) plan to their state or provincial energy regulator. These plans typically propose specific portfolios of programs that meet various policy objectives and cost-effectiveness guidelines and are usually devised for a one- to three-year cycle. Evaluation studies and other analyses are often timed to report in concert with the policy and budget approval processes, or even annually. In these cases, study findings that require follow-up or more research outside of the time frame of the policy/funding renewal cycle are unlikely to be included in evaluation plans.

Shareholder incentives and cost recovery mechanisms are generally based on *ex ante*, or pre-established, savings thresholds that trigger the compensation. Were participation levels anticipated in the program planning process met? There are exceptions, however, in which *ex post* evaluated savings estimates, based on evaluated savings parameters (e.g., lighting hours of use, washer cycles, net to gross ratios and so forth), are applied retrospectively. However, even these are generally based on first-year savings estimates and not considered over the lifecycle of measure savings.

Lost revenue recovery adjustment processes provide utilities with compensation for the revenues they do not receive because their energy efficiency programs reduce sales. The objective of these mechanisms is to decouple sales from profits so that utilities are not penalized for successful implementation of the efficiency programs. Determining the appropriate level of cost recovery requires knowledge of savings impacts over multiple years, including persistence, since utility rate-setting is generally based on future-year load forecasts.

### Energy Efficiency Resource Standards (EERS)

State EERS prescribe electric and gas energy and demand savings that must be achieved through energy efficiency programs and/or demand response programs. Most EERS targets were set five to ten years from the date of the legislative or regulatory enactment. Annual savings are usually cumulated, with EERS compliance being based on cumulative programmatic first-year savings over multiple years.

In determining compliance with EERS, most states seem to ignore questions of measure persistence. Without this adjustment, there is no way to compare utility compliance if one utility's program portfolio contains a large share of short-lived measures like CFLs to another that contains a large share of long-lived measures like building shell improvements. Without accounting for persistence, there is no way to compare the relative stringency and results of EERS between states.

Changing baselines could also affect compliance with EERS requirements. Narrowly focused first-year savings evaluations do not have to consider baselines in the second year after the measure is installed, much less the tenth.

### Benefit-Cost

As noted, all of the California SPM-derived benefit cost tests require estimates of lifetime savings of kWh and kW as a basic input to their cost effectiveness analysis. Thus, all tests need to be informed about likely savings persistence, which would imply the use of multi-year baselines for accuracy purposes. The TRC and the SCT specifically are further confounded by the lack of quality and timely data that is needed to estimate consumer incremental costs.

### Capacity Bids

Increasingly, wholesale electricity markets (e.g., PJM, ISO-New England) are allowing energy efficiency program savings to be bid into forward capacity auctions. Successful bids requires reliable and credible estimates of peak kW savings over the auction period.

It also draws attention to issues of multi-year savings persistence; currently, arbitrary and lowest-common-denominator expiration periods (e.g., 5 years regardless of the programs or measures) are imposed on energy efficiency program savings bids into the PJM system. Even though no such restrictions are in place for the ISO-New England system, documentation of persistence assumptions is required. More timely, accurate and measure-specific persistence data could help make a compelling case for a longer or more granular expiration schedules.

**Integrated Resource Planning Expectations**

System planners have to be skeptical about treating savings from energy efficiency as a long-term or reliable resource. Mid- and long-range system plans already have significant risk associated with "steel in the ground" generators due to construction delays, maintenance and unscheduled outage issues, unexpected fuel cost variations, and so forth. Even dispatch-friendly resources like large-scale demand response programs are difficult to allocate years into the future because, unlike a physical power plant, they could disappear due to "the stroke of a pen"—political and/or economic changes that are potentially more fitful than the usual variables that planners anticipate. Ironically, perhaps because few are really relying on these resources "as resources," they are more subject to political edict.

Traditional energy efficiency (that cannot be dispatched) is additionally subject to large question marks around its long-term impacts. Changes to codes and standards, prices, availability of new technologies, etc., are often difficult to predict accurately, yet can have major impacts on the difference between naturally-occurring conservation and that stimulated by energy efficiency programs. If the programs are to be taken seriously as system resources, long-term tracking of savings or at least estimation of survival curves is required. Both *ex ante* and evaluated assumptions need to be followed-up in subsequent years to confirm that useful life, performance and cost-effectiveness parameters have not strayed significantly from projections. This type of long-term study and analysis is almost non-existent in the energy efficiency program analysis world.

## Current Experience

As previous research reports[3] have illustrated, there is a broad range of evaluation and reporting practice across various North American jurisdictions when it comes to energy efficiency programs. Although good information exists about the most basic fundamentals of "who is doing what," there is a dearth of information about the specific details. In this case, the devil is in the details.

What we know, from an admittedly non-comprehensive examination of jurisdictions in in North America, is that it is very much the exception, not the rule, to find something other than first-year impact savings as the "currency" of energy efficiency results. Schiller and Goldman (2012) report that since most evaluation, measurement and verification (EM&V) studies focus on first-year savings, there is a lack of support for analyses of savings persistence. Comparative analysis of alternative program designs, estimates of market changes/effects, and mechanisms for prompt and regular program feedback are not emphasized.

As described above, much of the analysis that is actually needed is longer-term in nature. Lifetime savings of energy efficiency measures are important to understand and to factor into cost effectiveness analysis. Full life-cycle cost analysis (LCCA), which includes energy and costs involved in the manufacture, distribution and disposal of products in addition to purchase, usage and maintenance, would be even better.

Comprehensive research to answer more detailed questions like which jurisdictions use first-year impact estimates versus other options for official tabulations and reporting of energy efficiency savings would be helpful. Or, which jurisdictions may require other more granular information to be collected or reported like lifetime savings of the energy conservation measures; a full LCCA; multi-year baselines (to account for market and/or regulatory change); time-dependent avoided cost models (particularly for demand response), or; codes and standards analysis/evaluation?

[3] Several cross-jurisdictional comparisons are available, but particularly Kushler et al. 2012, Messenger et al. 2010 and a 2010 US DoE report (State Energy Efficiency Resource Standards Analysis).

Which jurisdictions essentially just repeat the first-year impact estimation process in successive years (therefore relying on a series of point estimates instead of true lifetime savings)? Which jurisdictions use a published TRM (or equivalent generated by the PA) savings values? Are these used only for *ex ante* purposes (program planning and projections) or are they used *ex post* (for savings calculations even when an evaluation may have produced alternate or updated assumptions)?

Which jurisdictions use EUL and RUL values in their TRM? Which jurisdictions have detailed incremental cost values in their TRM and update them regularly (often enough to be relevant)?

Even if we narrow the focus to only those jurisdictions that score well in the ACEEE 2012 State Energy Efficiency Scorecard[4] or the Canadian Energy Efficiency Alliance's National Energy Efficiency Report Card[5], there are few examples of leadership. For instance, only a few jurisdictions like British Columbia, California, Maryland, Massachusetts, Ontario and Oregon apply EM&V results retrospectively for savings goals and targets—meaning that the measured and evaluated results are the only results that count. (*Ex post* evaluation results are often used for cost-effectiveness assessment and calculations.) It seems probable that where official "accounting" of results is based on the retrospective, all involved would have a greater interest in the assumptions and processes for calculating them.

Even a top-ranked state like Vermont reports that "tracking system limitations require that annual measure savings remain constant for all years[6]." As for cost testing, there are almost as many variations in which tests are used and at what level (portfolio, program, or measure) as there are jurisdictions at the top of the aforementioned rankings. Long-standing and often cited sources like California's DEER (Database of Energy Efficient Resources) are expending effort to convert annual impact estimates into life cycle savings. They have had a contractor working on survival curves for various measures. In general, more experienced DEER users valued its comprehensiveness, with both savings and cost data for a wide array of measures. However, the data sources need to be more transparent and include date stamping (Jayaweera et al. 2011, p. 27).

Massachusetts is a model of thoroughness and transparency. It has a detailed Technical Resource Manual (TRM) of Prescriptive Measures. The current one, the TRM for 2013-2015, is reviewed and updated annually for estimated annual energy and peak demand impacts for primary and secondary energy sources. Their Planning & Reporting Information System (PARIS) includes updates from evaluation studies, according to an established update cycle and process, to plan for future program savings and to report actual savings.

PARIS is a statewide database maintained by the Department of Energy Resources (DOER) that emulates the cost effectiveness analysis tools of the PAs. As a repository for quantitative data from plans, preliminary reports, and final reports, PARIS generates information that includes funding sources, customer profiles, program participation, costs, savings, cost effectiveness and program impact factors from evaluation studies. Beginning with the 2010 plans, PARIS holds data from gas PAs as well. As of 2011, PARIS holds updates made to the PA tracking systems and cost effectiveness analysis tools to align with the assumptions and algorithms shown in the 2013-2015 Plan TRM.

Even Massachusetts' comprehensive exercise is for prescriptive (deemed) measures only. The Ontario Power Authority (OPA) publishes both fully documented prescriptive and quasi-prescriptive TRMs[7], including formulae and input variables for the quasi-prescriptive measures, and that is updated

[4] http://aceee.org/energy-efficiency-sector/state-policy/aceee-state-scorecard-ranking

[5] http://energyefficiency.org/wp-content/themes/ceea/ReportCard/2009%20Report%20Card_FINAL_lr.pdf

[6] Technical Reference User Manual Measure Cost Algorithms and Cost Assumptions; Efficiency Vermont; 1/1/2012.

[7] http://www.powerauthority.on.ca/evaluation-measurement-and-verification/measures-assumptions-lists

based on evaluated results of its own programs, but it doesn't have a PARIS-like transparent linkage to cost-effectiveness calculations.

In fact, in most jurisdictions, these are parallel endeavors—measuring program impacts and calculating ultimate cost effectiveness. Based on an informal survey of senior practitioners, one could strongly suggest that a significant disconnect exists between those who perform evaluations (mostly first-year) and those who perform cost effectiveness. Of the jurisdictions that use *ex post* evaluation results for official cost effectiveness calculations, unlike Massachusetts, many do not carry-through and ensure that the (hopefully) updated, locally-specific and current data informs prospective cost effectiveness calculations for future program designs and cycles. Of course, if enough successive iterations of good quality *ex ante* benefit-cost analyses exist, the end effect is almost the same, but many jurisdictions do not have a long history of evaluated results to rely upon (like California, for example).

## Why First-Year Savings Predominate

Limiting the scope of evaluations to first-year savings makes evaluations appear less costly and helps to avoid the need for up-front discussion of key uncertainties associated with the use of evaluation results in cost effectiveness analysis. But ultimately, this practice is a lost opportunity that simply transfers the costs and uncertainty of variables not collected in the first-year evaluations to the post-evaluation analyses. It often takes a few years before regulators and stakeholders become fully cognizant that multi-year savings estimates and incremental cost data are necessary to understand whether or not their policy objectives are being achieved.

***PAs*** want certainty in the form of operational clarity, especially because corporate and/or career rewards may depend on it. Since they are usually paid on the basis of results, PAs would prefer to focus on their (annual or short-term) delivery/sales missions and to not wait for a multi-year complex evaluation analysis to know whether they succeeded." Utilities or agencies that are responsible to pay PAs also more often prefer to have financial matters settled in an orderly and timely manner.

***Regulators*** are already faced with complaints from utilities and others that rate setting procedures are too onerous, too complex and time-consuming. Where lost revenue adjustment mechanisms are in place, there is already a bewildering array of methods to try to account for the level of conservation savings that are built-in to the load forecasts used for rate setting. EM&V results that contain uncertainty (which they do, by definition) add to the mix. All parties involved have a self-interested motivation towards simplicity and perhaps even over-simplification if it makes the "accounting" process shorter and less controversial. If first-year savings numbers are perceived to be more certain than other alternatives, the inertia quickly flows to using them and settling more quickly, especially when future reconciliation processes exist to deal with any resultant necessary rate adjustments.

Another pragmatic reality for PAs and regulators is that program schedules vary—not all programs start, stop or operate on a convenient annual cycle. The common way to deal with this is to "normalize" all savings to represent first-year savings for measures installed or actions taken through the program. The laborious accounting complexity involved in doing otherwise is usually perceived to outweigh the benefit of the granular-level of accuracy gained.

***EM&V practitioners*** themselves sometimes contribute to the "black box" perceptions about energy program evaluation. By tending to mask the true uncertainties inherent in evaluation results, or by focusing only on one or two elements of uncertainty[8], the whole enterprise is ultimately diminished.

[8] Many evaluators/evaluations report uncertainty based on sampling statistics, but do not report or include discussion of

These tendencies can be justified as pain-minimizing steps—clients often won't pay for more certain results; appropriate or enough data is not available or not made available; clients won't wait for more certain results; analyzing and then describing/communicating various levels of nested uncertainty can produce a serious headache for all involved, and so forth. It's easier to rely on aspects of measurement where there is engineering certainty rather than digging too deep into potential biases and study design issues that could call into question the credibility of the evaluation or the evaluator. First-year results may be "safer" in an environment somewhat closed to fulsome discussion and the open vetting of evaluation methodologies and results.

***Policy Makers***, perhaps more than any others, want certainty. They like to know far in advance if a program is "successful" or not—so that political capital can be expended to defend it, or, so that it can be contracted or eliminated in a controlled manner. Given the typical election cycle and how long it takes to design, implement and evaluate an energy efficiency program, the space in between is lucky to stretch to one year. Since the ultimate source of funding and approval usually comes from a political source, all other parties involved have a self-interest in feeding short-term, annual numbers into the "machine" for feedback purposes (positive or negative).

An emphasis on annualized numbers, accounting and feedback is hardly unique to energy program evaluation. Many acknowledge that longer-term results are important and worthy, but there's no convenient or established forum to deal with them. It is apparent from the private sector that even annual results are sometimes too slow. Focusing so much on quarter-to-quarter or annualized results has big risks for long-term sustainability, but it's now common practice in our society.

## Study Approaches that Can Better Meet Policy Objectives

Given the challenges described above, and because the energy program evaluation community has no clear path forward to address the issues associated with longer-term persistence and cost effectiveness, incremental steps will have to suffice while we await the revolution. We can contribute to broadening energy program evaluation horizons so they better encompass key questions and provide the answers that funders need to justify DSM programs. The solution is credible and reliable multi-year evaluation studies that account for variability of energy efficiency measure and program persistence, as well as estimation of impacts that would otherwise not occur due to market forces or other factors already included in reference load forecasts.

**Persistence** is the general term that deals with changes during measure life that are primarily due to retention and performance degradation. EUL is the term often used to describe persistence and is usually defined as an estimate of the number of years that 50% of the measures installed under a program are still in place and operable. EUL is not important for first-year savings estimates, but absolutely essential to understand what are the real cost-effectiveness and multi-year impacts. Regional-specific data can have a significant effect on the stream of savings where, for example, a rooftop A/C unit running almost daily in Arizona is unlikely to have the same EUL as one in Maine with a short cooling season, even though they came from the same manufacturer and are otherwise identical. For example, if average or median assumptions were used in Maine, or only first-year savings were the focus, A/C retrofit could appear to be not cost-effective when in reality it might be.

Clear identification of existing persistence studies and data that will be used to estimate multi-year savings should be an explicit part of energy program evaluation plans. Regional and national studies of this sort should be encouraged and expanded. Multi-jurisdiction studies could share the cost

---

measurement uncertainty, methodological uncertainty or the effect of potentially changing baselines over time.

and risk of this important research, making it more affordable and perhaps more useful, since it will have been widely-vetted.

**Dynamic Baselines** try to predict, based on some intelligent analysis, what future common practice will occur and factor its effect on baseline energy use, over periods of time. What is equipment life expectancy under real world conditions? How commonly is it left in place until break-down versus early replacement? What reasons is equipment replaced early—aesthetics, maintenance costs, energy savings due to more efficient and/or cheaper new models available, and so on? Are purchasing trends aligned with official codes and standards, or have they surpassed them?

Dual baselines, as described in an earlier section, are designed to explicitly change over the life of a measure, based on the estimated life of existing equipment and as standards change. A piece of equipment that can be retrofitted with a more efficient replacement has a RUL. If a program is premised on encouraging early replacement, then clearly the earlier that replacement occurs in the life cycle, the greater the impact. Equipment with one year of expected life left should not be credited (relative to the efficient replacement) with the same savings as equipment with 10 or 15 years of RUL. "Dynamic baselines" could incorporate more than two adjustment steps depending on changing market conditions and how long the measure lives continue.

Evaluations should answer the question of whether the market has been transformed (ahead of or because of codes and standards). Although code parameters are often used as a proxy for a baseline condition, any attempt to utilize a standard practice baseline instead of, or more likely in addition to, a project-specific baseline for the second and subsequent "steps" of the dynamic baseline would be a positive move. By using regional information on the behavior of cohort groups, evaluators can avoid relying completely on project-specific determinations, including the vagaries of self-report analyses of free ridership issues.

Dynamic baselines involve more work to create than single ones. Similarly, creating performance standard baselines is also more complex than just using site-specific information. Collecting data, developing models and estimating the variations in geography and customer/participant categorizations can be time-consuming and challenging—as can performing a large sample of project-based assessments. In the worst-case scenario, there may be little or no data or methods to interpret some cohort group behavior relative to devising a business-as-usual (BAU) case. Another important challenge is to avoid inadvertent overlap between net and gross savings at the step point (and how to handle the likely difference in attribution in the two periods). Despite these cautions, their use can be of significant benefit and cost-effective relative to using project-based baselines depending on the situation.

An expeditious approach has been developed for New York State (NYS) that reduces complexity and data collection effort and costs (Ridge et al. 2011). The concept is to use the ratio of incremental savings to full savings and the ratio of incremental costs to full costs to drive factors that NYS PAs can use to adjust the savings and cost data that are available. Tables were developed that only require the PA to have the RUL of the existing equipment and the full savings and costs of the new equipment or project (data that can be credibly produced by program staff). Although NYS-specific incremental cost information is not available, California's DEER was used since it contains the necessary detail. Differences in absolute costs and equipment performance between the two states certainly exist, but are not crucial if one believes that the ratios would be similar—the operating assumption in this case.

The NYS approach attempts to use <u>some</u> information to overcome a problem of <u>no</u> information. It is almost certain that baseline conditions will change in the real world, so assuming a null value (or no change) contributes to greater inaccuracy in long-term results. Ignoring change may also contribute to chasing savings that could ultimately end-up in the free ridership bin when a program continues too long to incentivize measures that end-up equivalent to "standard practice."

EM&V practitioners are generally forced to err on the side of conservatism. For example, in a retrofit program if clear evidence is not available to demonstrate early replacement, end-of-life will be assumed, but the opposite approach also occurs. Gathering data to better define baselines is therefore key to prevent the understatement or overstatement of savings, often due solely to a lack of information.

A **Multi-Year Approach to Attribution** recognizes that, like baseline conditions, program savings influences are dynamic, not static. Reliance on first-year savings estimates can lead to misinterpretation of Net-to-Gross (NTG) analysis.

What/who is a free rider or a free driver is rarely clear cut. Adoption of new technologies and behaviors is non-linear over time and therefore, not unexpectedly, so is free ridership[9]. Markets are dynamic and when interventions succeed, change is accelerated (that's the whole point of most energy efficiency programs). Nevertheless, it is very difficult to determine what a naturally-occurring "behavioral baseline" looks like when it may have been a decade or two since the first DSM programmatic interventions. Fixating on today's program, this year, may obscure the perspective that today's free riders may very well be yesterday's market effects; that those latent savings may not have been credited to previous generations of programs suggests future savings could be attributable to, but not credited to, today's programs. A multi-year approach to attribution (the recognition of dynamic NTG ratios) permits more intelligent program management over the life cycle of a program. Energy program evaluation can add considerable value to this process by providing evidence as to where along the market transformation "curve" the program currently sits. Providing that, and detecting any trends over multiple years, could be ultimately more valuable than providing a succession of "precise" in-year NTG ratios.

A **Broader Range and Depth of Econometric Analyses,** feasible due to an increasing availability of AMI data, should reduce costs for some types of multi-year evaluation. Some issues of persistence and behavior can be examined with this new rich dataset that could not practicably—or economically within most evaluation budgets—be looked at prior. More subtle load shape patterns across day types and seasons can be elucidated. The usefulness of these expands as time series data becomes available.

**Coordinating Evaluation with Potential Studies** is a help to system planners, but also contributes to the proverbial "closing the loop" evaluators constantly seek. In some cases, just doing a potential study would be a first step. Once one (or several) exists, evaluation plans can be coordinated to fill-in blanks in the potential study; findings can be triangulated amongst the potential study, traditional impact evaluation and market effects studies (should they exist). This "triangle" of complementary studies and datasets allows the monitoring, over time, of efficiency potential and whether it is being effectively tapped by energy program interventions. It also offers-up an ideal opportunity for PAs, evaluators and system planners to participate in a common (inter-agency) discourse about the goals and success rates of energy efficiency.

**Avoiding Exclusive Reliance on the TRC/SCT** until usefully accurate incremental cost data are available would also help foster a longer-term perspective. Several major energy efficiency organizations have criticized the application of the TRC/SCT in recent years, since most analyses include only selected costs and benefits. The concern is that the full cost to participants is almost always being counted while the full benefits (including, for example, increased comfort from HVAC or building shell measures) are not. This is in addition to the issues of dated and/or unavailable incremental cost

[9] A classic efficiency program targets a particular product in a particular jurisdiction in an attempt to overcome market and other barriers to the product. If promotion is successful, over an extended period, increased availability, higher consumer acceptance and lower prices for the product will cause the expected level of free riders to rise. Eventually, the product is mandated by a code or standard or becomes ubiquitous in the marketplace at which point every participant would be a free rider. Monitoring of and management of this process, using data from all of the studies discussed herein, allows for incentives to be appropriately reduced or withdrawn, over time, as the state of market transformation approaches.

data discussed earlier on and the apparent general lack of willingness to improve the situation. The tests themselves are not the problem, but the lack of care taken with their inputs may well be.

While a UCT leaves out even more costs and benefits, at least it is transparent and does not require incremental cost data that is so often unavailable (see also NYS example above). At a minimum, considering it alongside the TRC/SCT could be helpful, especially if the lack of good quality cost data embedded in the TRC/SCT is known and documented. Utilizing more, not fewer, of the California SPM Tests is generally a good policy along with recognizing the limitations of the quality of their inputs.

Also, cost effectiveness tools should bake-in the easy inclusion of impacts and costs by year, as appropriate. Our computers and software can now easily handle it and it’s not something individual projects or analysts should have to code and re-code. Simply prompting for these inputs could help broaden the awareness that they are important for the assessment of long-term resource savings and value.

Employing **Phased Research Plans** over the program cycle, or even longer, is probably the best way to ensure that all evaluation inputs are eventually addressed. Every study cannot practicably be completed every year, but a deliberate requirement to perform multi-year evaluation planning can help optimize limited evaluation resources so that the most impactful research is performed.

Many policy-derived program evaluation questions cannot be answered adequately within the confines of an annualized program approval and reporting cycle. An explicit multi-year phased evaluation plan helps decision-makers to recognize at what points enough data will be available to make higher confidence decisions about particular programs. It helps to erase the erroneous concept that all programs are one-size-fits-all—which can be exacerbated by the typical accounting-style portfolio reporting process that converts everything into a common, easy-to-tabulate, kWh, kW or BTU currency.

By staging and phasing various inter-connected studies, jointly funding more expensive and complex research, and applying some patience, significantly higher confidence results can emerge at the end of the process than had a succession of simple first-year savings studies been completed over the same period. Expansion of this concept to the portfolio level is even more impactful.

## Conclusion

An over-reliance on estimation of first-year savings rather than lifecycle savings creates lost opportunities in the policy world. This paper has described how this singular focus on first year saving could lead to significant errors in lifetime impact tracking and cost effectiveness analysis. Collecting and analyzing data related to persistence, baseline energy usage, and participant costs could be much more cost effectively done as part of an ongoing evaluation rather than as a separate activity, since much of the data could be obtained from program implementers and participants in conjunction with first-year evaluation data requests. In addition, the uncertainties associated with cost effectiveness analysis, compliance with EERS and needs of system planning are often obscured.

By focusing on reporting the higher precision achieved for key first-year savings parameters, rather than addressing the uncertainties in other key input parameters that were not collected, decision-makers could be provided inadequate and biased information that then leads to sub-optimal funding decisions and policy making. Taking a broader approach, using a wider array of study types over a longer time frame will yield results more appropriate for decision-makers to determine if policy objectives are being fulfilled by energy programs. This would help to increase the credibility of energy efficiency programs as energy resources and environmental mitigation measures, leading to more sustained commitment and reducing the likelihood of booms and busts in efficiency funding and activities.

## References


Allen, J.A. 2012. *Review of State Energy Efficiency Electricity Utility Program Savings Performance for 2010 – A Comparison of States with a Statewide Long-Term Savings Performance Target for Electric Utilities (Energy Efficiency Resource Standard or EERS) or Tailored Utility Targets, and States With Neither.* Montpelier, VT: Regulatory Assistance Project.

Barbose, G.L., Goldman, C.A., Hoffman, I.M. and Billingsley. M. 2013. *The Future of Utility Customer-Funded Energy Efficiency Programs in the United States: Projected Spending and Savings to 2025.* LBNL-5803E. Berkeley, CA: Ernest Orlando Lawrence Berkeley National Laboratory.

California Public Utilities Commission (CPUC). *California Standard Practice Manual: Economic Analysis of Demand-Side Programs and Projects.* October 2001.

Jayaweera, T., Haeri, H., Bruchs. D., Keeling, J., Khawaja, M.S., Rushton, J., Gowan, D., Carlson, S., Agnew, K., Goldberg, M., Jacobsen, D., Dimetrosky, S., Mort, D., Stern, F., Violette, D.M. and Baumgartner, R. 2013. *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures (January 2013 DRAFT).* NREL/SR-7A30-53827. Golden, CO: National Renewable Energy Laboratory.

Jayaweera, T., Haeri, H., Lee, A., Bergen, S., Kan, C., Velonis, A., Gurin, C., Visser, M., Grant, A. and Buckman, A. 2011. *Scoping Study to Evaluate Feasibility of National Databases for EM&V Documents and Measure Savings.* The Cadmus Group. Portland, OR: The State and Local Energy Efficiency Action Network.

Kushler, M.; Nowak, S. and Witte, P. 2012 *A National Survey of State Policies and Practices for the Evaluation of Ratepayer-Funded Energy Efficiency Programs*. Report Number U122. Washington, DC: American Council for an Energy Efficient Economy.

Lamont, D. and Gerhard, J. 2013. *The Treatment of Energy Efficiency in Integrated Resource Plans: A Review of Six State Practices.* Montpelier, VT: Regulatory Assistance Project.

Messenger, M., Bharvirkar, R., Golemboski, B., Goldman, C.A. and Schiller, S.R. 2010. *Review of Evaluation, Measurement and Verification Approaches Used to Estimate the Load Impacts and Effectiveness of Energy Efficiency Programs.* LBNL-3277E. Berkeley, CA: Ernest Orlando Lawrence Berkeley National Laboratory.

Ridge, R., Jacobs, P., Tress, H., Hall, N. and Evans, B. 2011. "One Solution to Capturing the Benefits of Early Replacement: When Approximately Correct Is Good Enough," *Proceedings of the 2011 International Energy Program Evaluation Conference*, Boston, MA: IEPEC.

Rufo, M. 2013. (Itron) Personal communication to authors. May, 2013.

Schiller, S.R. and Goldman, C.A. 2012. *Developing State and National Evaluation Infrastructures—Guidance for the Challenges and Opportunities of EM&V*. LBNL-4894E. Berkeley, CA: Ernest Orlando Lawrence Berkeley National Laboratory.

Schiller, S.R. 2011. "Documenting Greenhouse Gas Emission Reductions from End-Use Efficiency Activities in the Kyoto Protocol's Clean Development Mechanism Offset Program," *Proceedings of the 2011 International Energy Program Evaluation Conference*, Boston, MA: IEPEC.

Skumatz, L.A. 2011. "Remaining Useful Lifetimes and Persistence – Literature and Methods," *Proceedings of the 2011 International Energy Program Evaluation Conference*, Boston, MA: IEPEC.

Ting, M., Rufo, M., Messenger, M. and Loper, J. 2013 (forthcoming). *Measure Costs – The Forgotten Child of Energy Efficiency Analysis*. European Council for an Energy Efficient Economy 2013 Summer Study, Paper no. 7-389-13.