Ex-Post Verification and Cost-Effective Delivery of Energy-Efficiency Programs -- How to Produce Value without Breaking the Bank on Verification

Daniel M. Violette. Ph.D. Apex Analytics, Boulder, CO
e-mail: danv@apexanalyticsllc.com

ABSTRACT

The appropriate role of ex-post evaluations of energy efficiency (EE) programs has been debated on both sides of the Atlantic. This discussion often revolves around views that in-field verification is expensive, estimates of savings using ex-ante engineering-based approaches are good enough, and that monies would be better spent on delivering more EE measures and technologies rather than on expensive verification. In general, more resources are spent on ex-post verification in North America than in Europe. Most recently, this issue was elevated in discussions at the Fall 2019 International Energy Program Evaluation Conference (IEPEC) held in the United States where our evaluator colleagues from Europe questioned the need for and the value of verification approaches used in evaluations of EE programs in the United States. Similarly, when evaluators from North America travel to evaluation conferences in Europe and present papers on ex-post, in-field verification studies, we would often wonder at the lack of similar studies being conducted in Europe. This paper is meant to start bridging this discussion by presenting the underlying assumptions that might lead to a better understanding of the different views on in-field, primary-data verification studies, and the beliefs that result in the perceived value of these methods. Specifically, there may be practical issues that influence the selection of evaluation methods. In the U.S., large scale utility programs comprise many of the EE efforts. These programs may better lend themselves to ex-post verification methods. Also, utilities often face pressures to justify that the EE program expenditures made on behalf of ratepayers are prudently spent, i.e., they have to demonstrate that these expenditures on EE programs are producing the expected levels of energy savings.[1] Also, there is an established history of ex-post verification in North America that stretches back to the 1980s, with lessons learned regarding the value of in-field estimates of energy savings.

Introduction

This paper looks at how the application of ex-post monitoring and verification (M&V) methods in EE evaluation has developed in the United States. The use of in-field verification of savings from EE programs is more common in the U.S. as compared to Europe. This is, at least partially, a result of early experience with ex-post methods illustrating their value, the types of programs being implemented, and the regulatory environments in the United States. Factors that have influenced ex-post evaluation methods in the U.S. include:

[1] Investor-owned utilities (IOUs) in the U.S. often have regulatory/administrative proceedings around the design and review of EE programs. These proceedings involve stakeholders representing different customer and industry groups. Stakeholders can have disparate views on the amount of funds that should be spent and also on the EE programs to be implemented. In addition, some utilities receive financial incentives for reaching energy savings targets. Taken together, these factors can create pressures for the use of "high confidence" evaluation methods that may not be similarly present for programs in Europe.

- Learnings from early major evaluation efforts.
- Points of debate and critiques about the cost-effectiveness of EE programs that impacted the perceived need for ex-post evaluations.
- Issues in accuracy of engineering-based ex-post deemed savings.

Given these factors, two questions for current evaluation efforts are:
- Are ex-post verification studies cost-effective in today's environment?
- Can we conduct ex-post more cost-effectively that we do today?

This paper starts with some history of evaluation in the U.S., early lessons learned that still resonant today, and concludes with a discussion of whether the value of information produced by ex-post verification efforts exceeds its costs.

Early Ex-Post Evaluation (M&V) Efforts

Some of earliest large-scale evaluation efforts in the U.S. produced significant findings using ex-post verification regarding the actual value of savings from programs, and also on ways to make the programs more cost-effective going forward. In-field research on measure and program savings have produced important recommendations on how greater savings can be achieved through technology and implementation, or program redesign.

Two important, large-scale evaluation efforts dating back to the 1990s used ex-post evaluation M&V and set some precedents for using these methods. These are: 1) the National U.S. DOE Weatherization Assistance Program (WAP); and, 2) the Hood River Conservation Project (HRCP) in the Pacific Northwest are discussed.

Evaluation of the U.S. Weatherization Assistance Program (WAP)

An early path-breaking evaluation of the largest EE program in the United States helped jump start interest in comprehensive evaluation. The oil crises of the early 1970s had a significant impact on energy consumers. In response, Congress created the U.S. DOE's Weatherization Assistance Program (WAP) in 1976. By 1978, program expenditures hit $250 million per year. By 1985, approximately $2 billion had been spent on the DOE WAP program across the United States. Roughly another $2 billion was spent from 1986 to 1990. Expenditures peaked at nearly $500 million per year in 1988. Utilities often worked in concert with the U.S. DOE on the WAP, but utility contributions were less than 10% of expenditures over this 12-year period. From 1978 to 1990, the WAP represented the U.S.'s largest investment in energy efficiency.

Through about 1989, the program was prescriptive with a list of weatherization actions to be taken at each site. Air leakage controls using caulking around doors and windows were among the most common measures installed. Attic insulation was the next most common measure, with insulation in other areas (e.g., walls) occurring in about 20% of the installations. The program saw a number of enhancements over the 1978 to 1989 period with energy-efficiency improvements to windows and doors (e.g., storm door and window replacement). Additional EE measures tied to water heating and space heating were introduced by 1989. A more complete discussion of the WAP can be found in Brown et al. (1994).

Given the roughly $4 billion spent on this program from 1978 to 1990 (nominal dollars[2]), a number of review and audit actions were taken to ensure that the program was hitting targets. These targets were generally set in terms of counts of actions taken (e.g., number of homes/buildings treated and numbers of measures installed). The WAP was run through local and state agencies (e.g., state energy offices or community action agencies). From the start of the program, reviews of WAP activities across states were conducted for operational

[2] WAP expenditures in real dollars would be roughly 3X or 300% greater in 2019 dollars or $12 billion, so this was a significant national expenditure.

efficiency and to share best practices for efficient delivery of the program. What was missing in most cases was any actual ex-post measurement and verification (M&V) of attained reductions in energy use. This was due to:

1. A perceived high level of confidence that program measures were saving energy and that additional expenditures on in-field verifications were not judged as being needed; and,
2. A view that in-field evaluation efforts would divert funds from program implementation resulting in fewer dwellings being treated.

One of the first in-field assessments of national WAP savings was conducted by Ternes et al. (1988). This M&V study consisted of 66 occupied, low-income single-family homes in Madison, Wisconsin. Metering was used and ex-post verified energy savings were compared to the currently used pre-installation audits and ex-ante engineering calculation estimates. A second focus of the ex-post energy savings effort was on assessing new audit procedures and making recommendations for program delivery (e.g., the use of blower doors as part of the audit). Some finding from this ex-post evaluation are:

- The top-line finding was that in-field measured energy savings were considerably below predicted savings from the ex-ante pre-installation audit and engineering calculations – in some cases savings only 10% to 20% of predicted savings were actually being realized.
- The new, more costly audit procedures resulted in different measure installations being recommended when compared to current audit procedures.
- The use of blower door tests[3] and the enhanced audit protocols resulted in recommending measures that produced greater verified savings, but changes in energy use were still often inconsistent with the predicted savings.
- While enhanced audit protocols produced greater savings, they still did not accurately predict actual energy savings for individual houses. Savings were over-estimated with realized savings being only 50% of predicted in the tested homes. 20% to 60% of the deviation between predicted and measured savings could be attributed to incorrect assumptions regarding the indoor temperature before and after retrofit used in making the predictions. Occupants did not generally increase their indoor temperature after retrofit installation (the occupants did not generally display "take back" behaviour or a “rebound” in energy use).
- This resulted in a set of recommendations to the WAP. These recommendations included the use of blower door tests, enhanced audit protocols, and procedures that reduced the cost of installed measures by a factor of four. These significant cost reductions were achieved by selecting the most important measures in terms of predicted savings, rather than implementing a standard list of program measures at most all dwellings, whether they were needed or not.
- Overall, the new approach based on the ex-post analyses resulted in more than a doubling of the cost-effectiveness of the installed measures.

The final conclusion of the WAP ex-post evaluation was concise:

“In summary, the planning and implementation of a field test evaluation of building retrofits requires foresight, cooperation, and patience as different groups attempt to coordinate varied inputs to the project. It also requires money. The results, however, are well worth the effort, because the program becomes much more efficient in delivering effective energy conservation retrofits to low-income households.” (from Ternes et al., 1988 p. 47)

[3] The blowing door test in itself does not provide savings. Instead, the test is used to assess the before and after thermal integrity of the building and to identify sources of leakage.

A complete evaluation (Brown et al., 1994) of the WAP program covered installations made in 1989, 1990 and 1991. Over 350 local weatherization offices participated, data on approximately 15,000 installations were collected, and site-specific pre/post-energy consumption data were made available for 4,800 weatherized dwellings by regional utilities and state program administrators. By the 1989 program year, the use of blower doors were common and updated audit procedures were put into practice. This comprehensive WAP evaluation showed the overall program to be very cost-effective, but it also produced important recommendations for program improvements. The gap between ex-ante audit and engineering predicted savings, and actual ex-post evaluated savings impacted both EE program implementers/designers and evaluators.

The evaluation of WAP, along with the Hood River Conservation Project (discussed below), helped establish a view that taking a look to see what in-field, actual energy savings are for a program is important, even if it is only for small samples. Ex-ante audits and calculated savings estimates, even if grounded in good engineering, can still be problematic. The message from this work is that it is hard to anticipate all the differences that influence audit and engineering assumptions across participating installations and households. The in-situ conditions and site-to-site variability in conditions and behavior are often greater than what engineers anticipate when designing a program; and, in setting program implementation procedures and processes.

The lessons learned from this ex-post evaluation of this landmark national EE program in the U.S. raised a flag of caution that impacted the design of the next generation of program evaluations. Prudent evaluations now had to include not only activity metrics (e.g., number of measures installed and number of participants) and assessments of management/delivery/operational efficiencies; but, also a “look see” at the actual savings being achieved using in-field M&V on samples of actual participants. This “look see” does not have to be conducted every year, but frequently enough to provide confidence that the anticipated levels of savings were actually being attained in the field. Many organizations would do an early in-field program assessment to make early course corrections; then, perform in-field M&V updates every couple of years to manage costs.

Evaluation of the Hood River Conservation Project (HRCP)

The Hood River Conservation Project (HRCP) was a direct installation weatherization project implemented between 1983 and 1989 in Oregon’s Hood River and Wasco Counties. The Project was conceived by the Natural Resources Defense Council, which enlisted the cooperation of Pacific Power & Light Company, Bonneville Power Administration, the Hood River Electric Cooperative, the Northwest Power Planning Council, the Northwest Public Power Association, and the Pacific Northwest Utilities Conference Committee. All of these entities participated in a Regional Advisory Group, providing input and reaching consensus on decisions regarding project planning, implementation, evaluation and follow-up research. This produced a unique opportunity to work towards deep savings, collect detailed data to assess energy savings from different measures, and produce lessons learned for cost-effective programs. This was viewed as an evaluator’s dream project (See Hirst, E., 1988).

The HRCP was unique for its time with a detailed evaluation plan using pre- and post- data and participant surveys across three communities before the project began. The project had a continuing commitment to the collection and management of high-quality data, and to reporting not only its successes, but also its failures. It was intended to test “the reasonable upper limits of a residential retrofit program.” This three-year, $21 million research and demonstration project installed as many cost-justified retrofit measures in as many electrically heated homes as possible.

The evaluation used billing data in multivariate regression models to explain variations in monthly, seasonal and annual electricity use and savings across households. The models were designed to identify net savings (as well as total gross savings) attributable to HRCP. A number of advances in methods were made in this effort (Hirst, 1987 and White and Brown, 1990). The HRCP research documented new approaches for weather normalization, the use of control groups, and defining the pre- and post-participation time periods to account for lags between when the audit was conducted and measures implemented. The regression analyses also tested fixed effects and random effects specifications allowing for certain customer-specific factors to be captured by the regression. There was also a complementary end-use metering study for a smaller sample of homes.

Overall, deep savings were produced by the program and found to be cost-effective. Still, within these results, it was found that evaluated ex-post savings averaged only 43% of that predicted during energy audits. About 40% of the difference was attributed to discrepancies in the pre-HRCP energy use, and the remainder was caused by post-HRCP changes in energy-related behaviors (e.g., higher indoor temperatures and less use of wood). One confounding factor was a one-time jump in electricity prices in the period right before the program began. Also, the use of supplemental fuels (e.g., wood for heating) common to this region posed challenges as changes in the use of these fuels were not captured in the consumption analyses.

The fact that talented engineers relatively unconstrained by budget and methods were not able to predict ex-ante energy savings in line with the ex-post verified savings re-enforced the findings of the U.S. DOE WAP findings. This resulted in another caution to researchers – namely, don't just assume, verify using ex-post M&V. Like the WAP evaluation, many of the researchers in the HRCP study continued to break new ground in evaluation for decades. The acknowledgments page for the HRCP evaluation reads like a who's who of several decades of North American evaluation research.[4]

Other Early Evaluation Efforts

The DOE WAP program and the Hood River Conservation Program evaluations stand out in their scope and contribution. However, utilities were beginning to design and implement EE programs in the late 1980s and 1990s. Research entities such as the Electric Power Research Institute, Lawrence Berkeley National Laboratory, National Renewable Energy Laboratory (NREL) and others started energy efficiency and demand-side management programs. States such as California, Oregon, Washington, Wisconsin, Massachusetts (and New England), and New Jersey with its state-wide joint utility implementation of EE programs[5] were some of the leaders in program design, evaluation and cost-effectiveness analyses. Also, the International Energy Program Evaluation Conference (IEPEC) was founded in 1985 and became the clearinghouse for much of the leading evaluation work. The IEPEC has a searchable database of papers dating back to 1985, which is one of the best sources of practical insights into applied evaluation methods.

Evaluation Guidebooks/Protocols and Controversies/Scepticism

The early evaluation efforts jump started the industry going into the 1990s. Guidebooks for EE and DSM program evaluation were produced featuring guidance on ex-post verification along with examples. These include:

- Hirst, E., and J. Reed, eds. (1991). Handbook of Evaluation of Utility DSM Programs. Prepared for Oak Ridge National Laboratory ORNL/CON-336. https://www.osti.gov/servlets/purl/10120182
- Violette, D., M. Keneipp, M. Ozog, and F. Stern (1991), "Impact Evaluation of Demand-Side Management Programs: Volume I and II: A Guide to Current Practice." Electric Power Research Institute EPRI CU-7179, February.

---

[4] While I don't want to slight any of the participants in this effort, people that have influenced a generation of evaluators include Erik Hirst (the first winner of the life-time achievement award by the IEPEC), and others that have impacted our work include Meg Fels, Gil Peach, Ken Train (another life-time award winner), Marilyn Brown (lead researcher on the WAP evaluation), Ralph Cavanagh, Margie Gardner, Charles Goldman, Kenneth Keating (another life-time award winner), Martin Kushler, and Steve Braithwait.

[5] The New Jersey Conservation Analysis Team (NJCAT) project was another comprehensive effort where all seven of the State's electric and gas utilities worked together on the statewide delivery of a common set of programs. This included a cross-cutting multi-year, multi-million dollar evaluation effort (see RCG Hagler Bailly, 1990 and Ozog and Violette, 1990). The evaluation effort featured ex-post M&V for both residential and commercial/industrial programs.

- California DSM Measurement Advisory Committee (CADMAC), "Protocols and Procedures for The Verification of Costs, Benefits, And Shareholder Earnings from Demand-Side Management Programs," 1995 and 1998. *(Used to verify the achievement of goals required for utilities receive shareholder incentives from DSM programs)* http://www.calmac.org/events/PROTOCOL.pdf

Internationally, the IEA contracted for a streamlined evaluation guide:

- Violette, D. M. (1996), Evaluation, Verification, and Performance Measurement of Energy Efficiency Programmes, Prepared for the International Energy Agency, April 25.

The landscape for EE programs was changing in the 1990's as regulators in some states were encouraging utilities to aggressively pursue EE programs through shareholder incentives that made it in their interest to attain program goals. These incentives typically would be paid based on achieving verified EE goals. Requirements for levels of confidence and precision in evaluation results were established by state regulators, and protocols for measurement set out (see the CADMAC protocols above).

At the same time, there was a debate that included a number of noted academics about whether subsidized EE programs should be offered by utilities/states or whether the free market is a better way to allocate these resources. Joskow and Marron (1992) were among the leading voices arguing that utilities were not efficient at EE program delivery:

> ". . . *systematic biases in the reporting of costs and energy savings by utilities in our sample. In many cases utilities fail to report all relevant costs, rely on engineering projections of savings rather than applying methods to measure savings based on actual experience, and fail to make appropriate adjustments for free riders. Further biases may result from adopting measure lives that are too long. As a result, on average the cost of a (saved) nega-watthour computed from utility reports significantly underestimates the true societal cost of conservation achieved this way."*

It is notable that Joskow and Marron indicate that relying on engineering projections of savings are inappropriate and that it is important to apply methods that "measure savings based on actual experience, and adjustment savings for free riders." This pressure from academic researchers helped set out a level of rigor for reliable estimates of energy savings that included ex-post verification as a necessary component.

During this same period, some researchers were using top-down evaluation methods based on aggregate region-level data to estimate changes in energy use due to the aggregated expenditures across all the utilities located in that state. Several of these studies seemed to show that savings were lower than utilities' claims. A number of these top-down studies also produced estimates of EE program savings that were less than those produced by utility program evaluations. The best counter to the results of these top-down studies seemed to be the careful use of micro-data (i.e., customer-specific information), in ex-post verification studies where problems in terms of errors in variables, inconsistencies in how reported EE expenditures were aggregated across utilities in a State or Province, and the timing of those expenditures are not present. In addition, reviews of these top-down methods indicated that they had a number of challenges to overcome in producing reliable estimates of energy savings from utility or government-sponsored programs (See review of top-down methods in Violette and Rathbun, 2017 and in Violette et al., 2012).

The debates over the accuracy of EE cost-effectiveness analyses influenced the entire EE community including regulators, utility executives, state and federal government agencies with the result being that verification of energy savings became a high priority. This resulted in strong industry support for ex-post verification studies to meet the regulatory requirements of utilities for cost recovery and payment of incentives on successful EE programs.

The international community started to construct evaluation guidebooks at the end of the 1990s and into the 2000s that drew from work in North America, but with a different flavour designed to adapt methods to the types of programs offered in Europe. Examples of these are:

- Togeby, M. (2000): Use of Statistical Analyses in the Evaluation of Energy Efficiency Projects – an Introduction. København, AKF Forlaget.
- SRC International. "A European Ex-Post Evaluation Guidebook for DSM And EE Service Programmes," Prepared For: European Commission's Save Programme. April 2001. http://www.srci.dk/Files/Images/PDF/eu_ex_post_eval_guidebook_incl_appendixes.pdf
- IEA DSM (2005) Evaluating Energy Efficiency Policy Measures & DSM Programmes: Volume I Evaluation Guidebook; Prepared by Harry Vreuls, Operating Agent, Senter Novem, The Netherlands, October. http://www.ieadsm.org/wp/files/Exco%20File%20Library/Key%20Publications/Volume1Total.pdf

There are some common themes throughout these guidebooks:

- The need for good tracking systems with data collected at time of participation on both the status of the equipment being replaced to calculate baseline conditions and the operating profiles of the newly installed equipment – differences in baseline conditions across participants have been shown to be a significant component of variability in energy savings.
- As part of program tracking, develop the best initial estimate of savings that can then be used to make verification more cost-effective by allowing for appropriate stratification and efficient sampling.
- Plan the evaluation and M&V activities at the start of program implementation to allow for data to be collected at the time they are available, and put in place the algorithms used to estimate initial savings in the tracking data.
- Integrate evaluation efforts, i.e., ex-post M&V can be used not only to verify savings from past installations, but can also be used to make recommendations for program improvements both in the design and implementation of the program.
- Use ex-post M&V to promote and assess quality in the installation of all equipment -- poor quality installation and workmanship can result in savings leakage across the program. Even well-planned programs can suffer from poor execution.

Continuing Use of Ex-Post M&V in North America

The use of ex-post M&V in North America has continued to be a prominent feature of EE program evaluation. The lessons learned from the early evaluations of the WAP and the Hood River Project regarding the difficulties of anticipating the in-situ conditions for EE measures across a set of program participants have been borne out in more recent evaluation studies (see Korn and Dimetrosky, 2010 and Frankel and Turner, 2015). These studies not only have provided verified estimates of savings by program participants, but have led to numerous recommended improvements in program design.

Ex-post M&V for both residential and commercial lighting has been important (U.S. DOE Uniform Methods Protocols, 2017).[6] In the commercial sector, ex-ante engineering estimates have tended to over-state energy saving due to in-situ issues relating to estimating the hours of use, the number of lamps and fixtures installed (with some being put into storage by the building owner), and issues with a poorly specified baseline, i.e., baseline energy use being lower (or higher) than anticipated. Similar issues have turned up in M&V studies of residential lighting programs (Gaffney, K. et al. 2010) with hours of use and number of bulbs installed versus the number put into storage being different than assumed in ex-ante savings estimates. Turnover in building use or home ownership has also impacted persistence. These M&V studies have produced improvements in program design, marketing and delivery of lighting programs.

[6] Chapters 2 and 3 of the U.S. DOE Uniform Methods Protocols (2017) address the evaluation of commercial and industrial lighting and controls.

Examples where ex-post M&V produced energy savings that differed from the ex-ante estimates can be found for a number of EE measures and applications. Korn and Dimetrosky (2010) found verified savings from horizontal washing machines to be considerably lower than predicted. The M&V showed that, when energy use in dryers is also considered, then ex-post energy savings increased as clothes from a horizontal washer contained less water and need shorter drying times. This led to a recommendation for pairing horizontal washers with dryers equipped with functioning moisture sensors and make sure that users understand that energy savings are dependent on the use of automated drying cycles.

Frankel, M. and C. Turner (2008) examined 91 LEED-certified buildings and compared ex-post metered energy use to estimates based on ex-ante energy modeling. The ex-post metered energy use and savings varied considerably from the predicted ex-ante modeled values. Measured energy intensity for over half the projects deviated by more than 25% from the design projections, with 30% significantly better and 25% significantly worse. This wide range of accuracy between model-based predicted and ex-post measured adversely impacts building design decision-making, i.e., the assessment of alternate energy efficiency strategies based on the predicted energy savings and life-cycle cost analysis for each building. They concluded that "much work needs to be done to better align energy modeling accuracy with actual building performance outcome if this tool, as currently implemented, is to effectively serve the design community in delivering high performance buildings … and calls into question how effectively this tool is used to predict the performance outcome of any given project."

Some of the best sources for information on ex-post M&V studies are the papers in the Proceedings of the International Energy Program Evaluation Conference (IEPEC) which dates back to 1985, the American Council for an Energy Efficient Economy (ACEEE) Summer Study on Energy Efficiency in Buildings which dates back to 1982, and the U.S. DOE's Uniform Methods Protocols which addresses M&V for energy efficiency measures by application and sector.[7]

Refinement of Ex-ante Verification Method to Improve Cost-Effectiveness

There are clearly benefits that result from looking at the in-situ operational situations and energy use of program participants; however, there can be concerns about the costs of these ex-post studies. It is important that the costs of the research do not outweigh the value of the information to be produced (Violette and Moaz, 2019).

Researchers have worked to develop M&V approaches and methods that can help manage costs. These include:

- Use of ratio estimation approaches are now used by many, if not most, ex-post evaluation efforts. These approaches are designed to start with and leverage initial estimates of savings contained in a program tracking system. For this to work well, these estimates have to be participant specific and capture the participant-to-participant variability, i.e., if one participant is expected to have twice the savings of another, their initial savings estimate should be roughly twice as high. Developing ex-ante predicted savings values that capture as much of the participant-to-participant variability as possible allows for ratio estimation to achieve higher levels of precision with lower costs. This has become known as realization rate analyses and is the ratio of the realized to the predicted savings. If the initial predicted savings values are roughly accurate, then the realization rate (RR) ratio can be estimated more precisely than are the standalone site energy savings levels.[8]

---

[7] The DOE UMP can be accessed at https://www.energy.gov/eere/about-us/ump-home; the IEPEC proceedings can be found at https://www.iepec.org/ using the search proceedings tab; and the ACEEE Building and Energy Use conference proceedings can be found at https://www.aceee.org/proceedings.

[8] Program savings are estimated by multiplying the overall program ex-post realization rate times the initial program-wide savings estimate. A good example of realization rate estimation is in Violette and Rogers (2014) and in the section on "Paired Data to Enhance Precision" in Hirst and Reed (1991) in *Chapter 4: Analyzing data* by Violette, D. M.

- Investing in high quality tracking systems can reduce the overall costs of evaluation. Protocols for tracking systems can be just as important as protocols for M&V.[9] There is only one chance to collect data on the baseline equipment being replaced and on many of the baseline operational variables. Once the new EE equipment has been installed, it may be impossible to go back and get information on the efficiency of the equipment replaced that is needed for accurate baseline calculations. As a result, this baseline data must be captured by the program tracking system. This has been a significant limitation for ex-post M&V as you need both pre- and post-installation energy use to calculate savings. Often, the variability in baseline conditions across participating sites is greater than anticipated causing unexpected variation in site-to-site savings estimates. In addition, clear information is needed on where different measures have been installed, when they were installed, and the appropriate contact information for people that can address questions from evaluators.
- Program tracking systems are the source of information needed for drawing cost effective sampling that includes stratification by energy use, and often by end-use or building types (e.g., office buildings, schools, or retail). A discussion of how cost-effective samples using ratio estimation can be developed is contained in Violette and Rogers (2014).
- Integrate M&V with program implementation so that customers can be randomly drawn as they participate in the program. This allows for pre- and post-installation metering and data collection on baseline conditions. Often a rolling sample process is used where, for example, every 7th participant[10] is placed into the ex-post M&V sample. This can also help produce near real-time findings on savings as the program is being rolled out which can help deliver timely recommended improvements to the program design and implementation.

Budgeting for Ex-Post M&V

Evaluation plans and budgets that provide funding for ex-post M&V have also evolved over time. Now, most evaluation budgeting processes in North America involve planning over a 3-year to 5-year time period, and a budget that is developed for a portfolio of programs. These portfolios typically have programs in several sectors (e.g., residential, commercial, large commercial & industrial, and low income) and may comprise between 8 to 12 programs. The Consortium for Energy Efficiency (CEE) tracks evaluation budgets for each state in the U.S. and each Province in Canada for electric and gas programs. The 2018 CEE Annual Industry Report estimates total evaluation budgets for the U.S. to average 2% of DSM Expenditures (EE and DR), and at 4% for Canada. These budgets are comprehensive as they not only include impact evaluation, but also process evaluation and market research. When you look at the state-by-state breakdown in the U.S., states that would be viewed as leaders in DSM investment tend to have slightly higher evaluation budgets that are closer to 3% -- possibly indicating their perceived value of evaluation (broadly defined) in delivering high-quality programs.

Expenditures of 3% of total program dollars across impact, process and R&D does not seem extravagant even with the amount of ex-post M&V that is conducted in North America. It would not take much for the evaluation efforts to pay for themselves in terms of recommended improvements in program design, new program development, and implementation/delivery. The WAP evaluation (Brown et al., 1994) and the Hood River Project (Hirst, 1987) increased the cost-effectiveness of the program by more than a factor of 2 over initial designs that were based on ex-ante engineering and audit-based estimates. The allocation of these dollars to

[9] Many of the companies that are used as 3rd party implementors in North America advertise the comprehensiveness of their program tracking software. As just one example, CLEAResults (https://www.clearesult.com/solutions/) has DSMTracker™, but all the major vendors have program tracking software to manage programs and provide data for evaluation. DSMTracker is just one of these offerings.

[10] The use of every 7th participant to comprise sample (i.e., you randomly pick a starting point number) is not based on any statistical rule of thumb. It is just an example and the number chosen should reflect the projected participant population and the expected variance of the energy savings values.

different program evaluation activities takes into account the timing and value of the research. For example, ex-post M&V involving primary field data collection is not viewed to be necessary every year. Ex-post studies on major programs are usually conducted early in program implementation to capture initial improvements/recommendations and then roughly every 3 years, or after major changes to programs. In addition, there is often an effort to focus the ex-post effort on segments within a program that are believed to be most uncertain and in need of verification. This uncertainty may be based on research conducted in other regions or by techniques such as error propagation methods used to assess uncertainties in engineering algorithms or models.

Economies of scale in evaluation are also important. A program that reaches 5,000 customers may not be much more expensive to evaluate than a program that reaches 500 customers. This is due, in part, to the sample size requirements, i.e., a program that is twice as large does not need a sample that is also twice as large. As a result, states and regions in North America often work together on evaluation efforts to help reduce costs. California performs state-wide evaluation spanning programs delivered by all the state's utilities to help achieve economies of scale. This has become common place for most states where a state-wide evaluation is commissioned to lead and manage the evaluation work. Regional studies also take place in some areas. For example, the Northwestern states of Washington, Oregon, and Idaho often pool evaluation resources.

In addition to the CEE Annual Reports on DSM program and evaluation expenditures, there are a number of good references on evaluation planning and budgeting that assess the tradeoffs in methods and costs. SeeAction (2012) has a section on budgeting from the perspective of the value of evaluation results and uncertainty. The IEA DSM Evaluation Guidebook (2005) also sets out principles for deciding on evaluation activities and levels of effort that are similar to those set out in the more recent SeeAction report.

Conclusions

There is clearly a difference in priorities when it comes to funding and conducting ex-post M&V on energy efficiency programs between North America and Europe. This has been demonstrated in the published literature and in conference proceedings, as well as comments by evaluators at major conferences. The reasons for the difference in approach seems to have little to do with professionalism or a desire for accurate information on actual energy savings. Instead, it appears to be a difference in point of view, i.e., on the perceived value of conducting ex-post M&V.

Factors influencing the North American view on the value of ex-post M&V includes:

1. A history of M&V dating back to evaluations of large-scale national EE programs where in-field data collection on energy savings had been delayed for a number of years. Initial evaluations were based on activity metrics such as the number of participants, EE measures installed (with deemed savings estimates), funds committed, and quality metrics based on training of auditors. However, when ex-post M&V was finally conducted, the verified energy savings from these national programs were found to be only a small fraction of expected/predicted savings. In some cases, verified savings for program segments were only 10% to 20% of predicted savings. This implied that the U.S. had been implementing EE programs for a number of years that did not meet the cost-effectiveness criteria set out for these programs.
2. The lessons learned from the history in item #1 above indicated that predictions of energy savings based on ex-ante engineering calculations and energy audits were not as robust or as reliable as believed. The in-situ diversity in behavioral and operational conditions across a large sample of actual program participants was greater than what was currently assumed resulting in large differences between predicted and measured energy savings.

3. The evolution of evaluation methods in North America was driven by the professionals working on these early ex-post evaluations. These professionals became the backbone of evaluation in the U.S. and Canada. The lessons learned from these early ex-post evaluations influenced a generation of evaluators.
4. There was an intense debate in the 1990s in North America over whether EE should be delivered through utility and governmental programs, or left to free markets. This pressured utility and other program administrators (e.g., state energy offices) to seek high levels of rigor for estimates of energy savings to confirm cost-effective program accomplishments. This resulted in the use of ex-post M&V approaches to deliver energy savings estimates that met a high level of reliability.
5. Also, during the 1990s, regulators of utilities at the state level considered, and many implemented, financial incentives to achieve high levels of energy savings through programmatic DSM. With monies at stake, regulators required estimates of EE program accomplishments to be estimated at high levels of confidence and precision.
6. Given the pressures of items #1 through #5 above, the evaluation community in North America worked to develop ex-post evaluation methods that were based on leveraged ratio estimation, rigorous tracking systems to collect data from program participants at the time of participation, and the development of sampling protocols that would help manage the costs of ex-post M&V.
7. Budgeting for evaluations of portfolios of EE programs evolved to consider the timing of ex-post M&V within 3-year and 5-year planning horizons. These multi-year portfolio evaluations allow for ex-post M&V to be conducted at different points in time to help manage costs of in-field data collection. This helps produce evaluation plans for large-scale programs that incorporate ex-post M&V within budgets that were approximately 2% to 3% of total program expenditures. These evaluation budgets included process evaluation, market research, as well as impact evaluations with ex-post M&V on the largest (or the most uncertain) programs.
8. The experience with ex-post M&V demonstrated that it not only produced reliable estimates of historical energy savings, but it could be instrumental in the development recommendations tied to program design and delivery/implementation. These recommendations improved program cost-effectiveness by more than the cost of evaluation.
9. Ex-post M&V supported and provided checks on the quality of delivery and implementation. Over the years, there have been a number of quality issues observed related to EE program implementation. This cost-effectiveness leakage due to a lack of quality could be as high as 10% to 20% of program expenditures in some cases.[11] Promoting quality of program delivery and implementation maybe be the most under-rated aspect of ex-post M&V. It serves as a real-world audit of program accomplishments and provides a check on the quality of program management, implementation, and installation of EE measures. It helps provide for accountability for program performance.

The factors above set out beliefs and values used to support the use of ex-post M&V methods in assessing portfolios of EE programs in North America. Two major counter arguments for the use of in-field data collection to support energy savings estimation are:

1. Current energy savings estimates based on ex-post engineering algorithms, audits and building modeling are accurate enough, and the value of additional ex-post M&V does not outweigh the costs.
2. Evaluations based on activity metrics such as numbers of participant sites and measures installed are adequate for assessing program progress towards goals, and can be collected at lower cost than ex-post

[11] Quality issues ranged from simple non-performance by contractors, to poor installation work, to a variety of shortcuts that could reduce costs but would also lower the likelihood of achieving expected energy savings. The fact that participant sites were to be visited by M&V personnel to assess program delivery, installation and energy savings helped provide a quality check across program activities.

M&V. These metrics also provide information on cost-effectiveness of program delivery and the insights required for recommending improvements to the program; and, next generation program design.

The relative merits of the supporting arguments for ex-post verification of energy savings versus the of energy counter arguments need to be assessed against the goals of a program and the evaluation effort needed to show those goals are being achieved. It may be a matter of belief in the accuracy of different methods (e.g., ex ante engineering estimates are accurate enough) and the experiences of the evaluation community.

As a final note, both the United Kingdom (UK) and Germany are currently offering aggressive commercial sector audit programs. In-field data collection on measures installed through audit programs along with measured energy savings have been completed for similarly complex programs across North America.

The UK has the Energy Savings Opportunity Scheme (ESOS) which requires all large enterprises (or smaller organisations that are part of a large undertaking) to carry out audits or a specified equivalent or exemption (such as ISO 50001) of the energy used by their buildings, industrial processes, and transport to identify cost-effective energy saving measures. A process evaluation was conducted in 2017 that addressed a number of delivery and market transformation issues (United Kingdom, 2017). The next step is an impact evaluation of the ESOS. An ex-post M&V philosophy would look at actual measures installed with measured energy savings metrics. This would involve close cooperation between the evaluators and program implementation administrators to provide for the timely collections of program data that would facilitate a quality ex-post energy savings evaluation at a reasonable cost.

Germany (2019) has the "SME Initiative for Energy Reforms and Climate Protection." This initiative provides federal funding for energy audits that provide companies with specific suggestions for commercially rational measures to boost energy efficiency. Germany is also offering "federal funding for energy efficiency in commerce" program as of 2019. Ex-post evaluation of these custom measure programs poses challenges but, by collaborating with the program implementors/administrator, ex-post methods can provide energy savings based on in-field, primary data. These impact evaluation challenges have been met in other large-scale evaluation efforts on important programs for achieving climate goals. Ex-post evaluations not only provide estimates of actual attained energy savings, but they also provide the real-world feedback needed to improve the overall quality of energy efficiency programs by addressing inadvertent and unforeseen leakages in energy savings.

References

Brown, M.A., L.G. Berry and L.F. Kinney 1994. "Weatherization Works: Final Report of the National Weatherization Evaluation." Prepared for the Weatherization Assistance Programs Division U. S. Department of Energy, Report # ORNL/CON-395, September. https://www.osti.gov/biblio/366506

CALMAC 1998. "Quality Assurance Guidelines for Statistical, Engineering, and Self-Report Methods for Estimating DSM Program Impacts." Prepared for the California Demand Side Management Advisory Committee, By Rick Ridge, Dan Violette and Don Dohrman, July. http://www.calmac.org/events/APX_J%200698.pdf

Consortium for Energy Efficiency - CEE 2018. "Annual Industry Report" https://library.cee1.org/content/2018-cee-annual-industry-report/

Frankel, M. and C. Turner 2008. "How Accurate is Energy Modeling in the Market?" *Proceedings of the 2008 ACEEE Summer Study on Energy Efficiency in Buildings*, Asolimar, CA p. 3-88. https://www.aceee.org/files/proceedings/2008/data/papers/3_320.pdf#page=1

Gaffney, K. et al. 2010. “I Know What You Lit Last Summer: Results from California’s Residential Lighting Metering Study, *Proceedings of the 2010 ACEEE Summer Study on Energy Efficiency in Buildings* p.9-80. https://www.aceee.org/files/proceedings/2010/data/papers/2225.pdf

Germany Federal Ministry for Economic Affairs and Energy 2019. “Energy Consulting and Funding for Companies,” https://www.bmwi.de/Redaktion/EN/Artikel/Energy/energy-consulting-and-funding-for-companies.html

Hirst, E. 1988. “The Hood River Conservation Project: An Evaluator's Dream.” *Evaluation Review*. Volume: 12 issue: 3, page(s): 310-325 https://journals.sagepub.com/doi/abs/10.1177/0193841X8801200306

Hirst, E., D. L. White, and R. Goeltz 1987. “Three Years after Participation: Electricity Savings Due to the BPA Residential Weatherization Pilot.” Research sponsored by the Bonneville Power Administration, U.S. Department of Energy and the Pacific Power & Light Company. Report #: ORNL/CON-231. https://www.osti.gov/servlets/purl/6880640

Hirst, E., and J. Reed, eds. 1991. in *Handbook of Evaluation of Utility DSM Programs.* Prepared for Oak Ridge National Laboratory, ORNL/CON-336. https://www.osti.gov/servlets/purl/10120182

IEA DSM 2005. *Evaluating Energy Efficiency Policy Measures & DSM Programmes: Volume I Evaluation Guidebook*; Prepared by Harry Vreuls, Operating Agent, Senter Novem, The Netherlands, October. http://www.ieadsm.org/wp/files/Exco%20File%20Library/Key%20Publications/Volume1Total.pdf

Joskow, P. and D. Marron 1992. “What Does a Negawatt Really Cost? Evidence from Utility Conservation Programs.” *The Energy Journal* Vol. 13, No. 4. https://www.jstor.org/stable/41322467?seq=1

Korn, D. and S. Dimetrosky 2010. “Do the Savings Come Out in the Wash? A Large-Scale Study of In-Situ Residential Laundry Systems.” *Proceedings of the 2010 ACEEE Summer Study on Energy Use in Buildings*, Asolimar, CA p.9-143. https://www.aceee.org/files/proceedings/2010/start.htm

McCold, L.N. 1987. Field Test Evaluation of Conservation Retrofits of Low-Income, Single-Family Buildings: Combined Building Shell and Heating System Retrofit Audit. Oak Ridge National Laboratory. Report # ORNL/CON-228/P3. https://info.ornl.gov/sites/publications/Files/Pub57671.pdf

Ozog, M. and D. Violette 1990. Using Billing Data to Estimate Energy Savings: Specifications of Energy Savings Models, Self-Selection and Free-Riders. *Proceedings of the American Council for and Energy-Efficient Economy (ACEEE) 6th Summer Conference on Energy Efficiency in Buildings.* https://www.aceee.org/files/proceedings/1990/data/papers/SS90_Panel6_Paper15.pdf

RCG/Hagler Bailly, Inc. 1990. "New Jersey Conservation Analysis Project: Final Contractor's Report to the NJCAT," prepared for The New Jersey Conservation Analysis Team, August 14.

SEE Action 2012. “Energy Efficiency Program Impact Evaluation Guide.” Prepared by Steven R. Schiller, Schiller Consulting, Inc., https://www4.eere.energy.gov/seeaction/evaluation-measurement-and-verification-resource-portal

SRC International 2001. “A European Ex-Post Evaluation Guidebook for DSM And EE Service Programmes,” Prepared For: European Commission’s Save Programme. April. http://www.srci.dk/Files/Images/PDF/eu_ex_post_eval_guidebook_incl_appendixes.pdf

Togeby, M. 2000. Use of Statistical Analyses in the Evaluation of Energy Efficiency Projects – an Introduction. København, AKF Forlaget.

United Kingdom Department for Energy and Climate Change – DECC 2017. "Evaluation of the Energy Savings Opportunity Scheme: Interim process and early impact evaluation report." Prepared by Ipsos MORI and University College London (UCL). October. https://www.gov.uk/government/publications/energy-savings-opportunity-scheme-esos-evaluation-of-the-scheme

U.S. DOE Uniform Methods Protocols 2017. "The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures." National Renewables Energy Laboratory, NREL/SR-7A30-53827 April 2017. https://www.energy.gov/eere/about-us/ump-home

Violette, D. et al. 1991. *Impact Evaluation of Demand-Side Management Programs, Volume I and II: A Guide to Current Practice*. Electric Power Research Institute. February, 1991. EPRI CU-7179.

Violette, D. 1995. Evaluation, Verification, and Performance Measurement of Energy Efficiency Programmes. Prepared for International Energy Agency.

Violette, D. et al. 2012. "Bottom-Up and Top-Down Approaches for Assessing DSM Programs and Efforts." *Proceedings of the 2012 International Energy Program Evaluation Conference*, Asolimar, CA, August. https://www.iepec.org/wp-content/uploads/2018/02/080-7.pdf

Violette, D. and B. Rogers 2014. "A Sampling Methodology for Custom C&I Programs." Prepared for Technical Evaluation Committee, Ontario Energy Board sponsored by Union Gas and Enbridge, October. https://www.researchgate.net/publication/342165724_A_Sampling_Methodology_for_Custom_CI_Programs_Technical_Evaluation_Committee_--_Updated_October_2014

Violette, D. and P. Rathbun 2017. Chapter 21: Estimating Net Savings – Common Practices. *Methods for Determining Energy-Efficiency Savings for Specific Measures*. Golden, CO; National Renewable Energy Laboratory. NREL/SR-7A40-68578. http://www.nrel.gov/docs/fy17osti/68578.pdf

Violette, D. and K. Moaz 2019. "Estimating Savings Persistence – Value of Information and Continuous Learning Approaches." *Proceedings of the International Energy Program Evaluation Conference,* Asilomar, CA. https://www.iepec.org/wp-content/uploads/2020/03/Violette-Daniel-Estimating-Savings-Persistence-%E2%80%93-Value-of-Information-and-Continuous-Learning-Approaches.pdf

White, D. L. and M. A. Brown 1990. "Electricity Savings Among Participants Three Years after Weatherization in Bonneville's 1986 Residential Weatherization Program," ORNL/CON-305, Oak Ridge, TN: Oak Ridge National Laboratory, September. https://www.osti.gov/servlets/purl/6454513