Lessons Learned in Selecting Impact Evaluation Methodologies: Metering, Modeling, or Billing Analysis?

Robert Kasman, Pacific Gas & Electric Company[1]

ABSTRACT

What issues should be considered when selecting the best impact evaluation method for demand-side management (DSM) program evaluation?

Although significant literature is available on the details of *implementing* particular evaluation methodologies, much less is available on *selecting* the best evaluation methodology. Due to the specifics of each DSM program and its associated evaluation requirements, a definitive algorithm for optimal methodology selection does not exist – experience is the guide.

Utilizing the recently published evaluation of the 2004-2005 California Energy Star® New Homes Program, this paper contributes to ideas in the selection of methodologies for impact evaluation. The Energy Star Homes evaluation is ideal for this purpose since it utilized three different evaluation methods: end-use metering, whole building modeling and billing analysis. The approach and advantages of each evaluation methodology are summarized, along with key challenges, usefulness of results, practical challenges, and recommendations to overcome challenges.

This paper is targeted to both program evaluators and implementers to enhance future evaluations and usefulness of their findings.

Introduction

The purpose of this paper is to aid in the selection of impact evaluation methodologies for *all* DSM program types. Although it uses examples from the 2004-2005 California Energy Star New Homes Program evaluation, this paper is *not* particularly about residential new construction, and *not* about the Energy Star Homes program.

Simplified project-level evaluation methodology selection guidelines do exist (EVO 2007, 37), as well as specific examples (EVO 2007, 66-84), but these do not address the complex or practical challenges associated with large DSM program impact evaluation planning.

Background

The California Energy Star® New Homes Program provides financial incentives, education, and marketing assistance to California builders who construct new residences that exceed the state's mandatory minimum energy efficiency standards, Title 24.[2] Builders must construct new homes at least 15% better than Title 24 to participate in the program.

---

[1] Disclaimer: The views and opinions expressed in this article are solely those of the author and are not intended to represent the views and opinions of Pacific Gas & Electric Company.

[2] http://www.energy.ca.gov/title24/

Identifying Key Evaluation Challenges

Every evaluation faces challenges – identifying these early during evaluation planning is critical to sound methodology selection. In the case of the Energy Star Homes evaluation, there are three key challenges: there is no easily identifiable baseline (there is no "pre" condition), isolation of energy conservation measures (ECMs)[3] are not possible, and there is usually huge variance in occupant behavior. Note that other program types often share some or all of these challenges.

No Easy Baseline

When estimating energy impacts, establishing the baseline is half the challenge.[4] In new construction impact evaluation, determining baseline energy usage is inherently difficult since there is no pre condition as in retrofit projects, and there is no identical non-participant[5] control structure to compare to. Therefore, even with detailed participant structure consumption data, there is no easily tenable baseline to compare to. This is a key challenge of new construction impact evaluation.

Isolation of ECMs

Isolation of ECMs is generally not possible in new construction. Design changes can include "measures" such as infiltration sealing and testing, cool roofs, radiant barrier, window type, window area, and many others. Due to the installation of multiple measures at a single site, and interactions between these measures that affect energy use, isolating these "measures" and/or directly metering their energy consumption is not possible.

Occupant Behavior

Occupant behavior can vary dramatically; example shown in Figure 1. This is critical to recognize for all analysis methods using metering or billing analysis since the variation in occupant behavior is "noise" to the "signal" that is being inspected. Anticipating occupant behavioral variance is a key factor for consideration in selecting an evaluation methodology, and is also applicable to many non new construction evaluations. As the behavioral variance increases, the usefulness of metering and billing analysis decreases. How great can behavioral variance be? Parker, Mazzara, and Sherwin 1996, reported,

> *The variation of space conditioning needs arising from occupant behavior has been consistently noted in previous monitoring efforts. Early studies at Princeton's Twin Rivers project showed differences between otherwise identical*

---

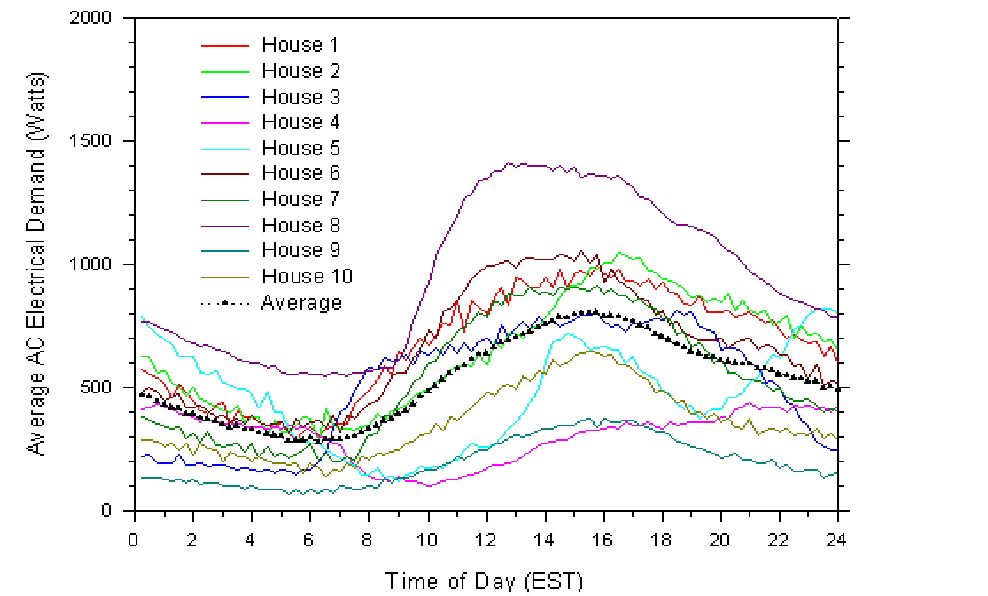
[3] ECM is the terminology used by International Performance Measurement and Verification Protocol (IPMVP), however the DSM community in the United States uses the term conservation to refer to behavioral changes resulting in energy and demand savings, in contrast to energy efficiency measures, which are technologies that increase energy (and/or demand) savings without user action, or knowledge, required.

[4] The basic equation of energy savings is: Savings = (Baseline-period use or demand) – (Reporting-period use or demand) +/- adjustments

[5] Participants and non-participants are used in the context of energy efficiency program participation, or not, and the resulting efficient, or baseline efficient, buildings.

townhouses of 2:1 in space conditioning energy (Sonderegger, 1978). Similarly, a study of air conditioning use in 25 homes in Palm Beach, Florida showed a 100:1 variation in space cooling energy, mainly based on differences in ventilation behavior (Parker, 1990). The variation was still 7:1 when primarily air conditioning households were considered.

Figure 1: Example of Typical Occupant Behavioral Showing Large Variance

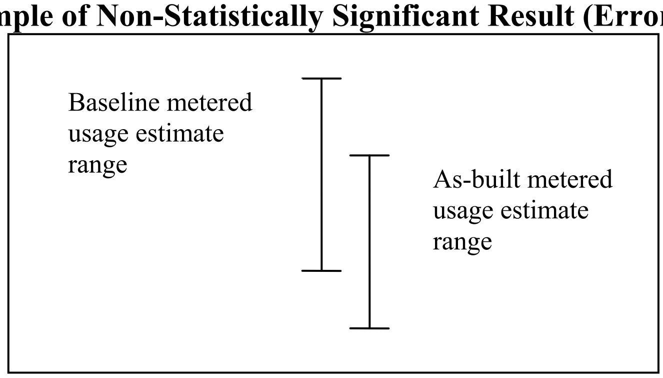


24 Hour AC load profile for all ten homes and average.

Source: (Parker, Mazzara, and Sherwin 1996)

The effect of the large behavioral differences is to create large uncertainty around the mean usage. Unless the baseline usage estimate is significantly larger, the uncertainties will swamp the difference of the means, leading to inconclusive results, shown in Figure 2 below.

Figure 2: Example of Non-Statistically Significant Result (Error Bars Overlap)



This common result will likely be reported as, "…the study showed an indication of savings, but the results were not statistically significant, and additional research is needed to…" One goal of sound evaluation design is to avoid this outcome when possible.

Consider Standard Project Evaluation Methods

IPMVP (EVO 2007) are widely referenced for conducting M&V activities at the project level, with program level impacts determined through appropriate sampling and extrapolation to population.[6] Table 1 shows a summary of the methods applicable to new construction projects.

Table 1: IPMVP Overview of New Construction M&V Options

M&V Option	How Baseline is Determined	Typical Applications
A. Partially Measured ECM Isolation Savings are determined by partial measurement of the energy use of the system(s) to which an ECM was applied, separate from the energy use of the rest of the facility. Some parameters are stipulated rather than measured.	Projected baseline energy use is determined by calculating the hypothetical energy performance of the baseline system under operating conditions during the M&V period.	Lighting system where power draw is periodically measured. Operating hours are stipulated.
B. ECM Isolation Savings are determined by full measurement of the energy use and operating parameters of the system(s) to which an ECM was applied, separate from the rest of the facility.	Projected baseline energy use is determined by calculating the hypothetical energy performance of the baseline system under measured conditions during the M&V period.	Variable speed control of a fan motor. Electricity use is measured on a continuous basis throughout the M&V period.
C. Whole Building Comparison Savings are determined at the whole-building level by measuring energy use at main meters or with aggregated sub-meters.	Projected baseline energy use determined by measuring the whole-building energy use of similar buildings without the ECMs.	New buildings with energy-efficient features are added to a commercial park consisting of buildings of similar type and occupancy.
D. Whole Building Calibrated Simulation Savings are determined at the whole-building level by measuring energy use at main meters or sub-meters, or using whole-building simulation calibrated to measured energy use data.	Projected baseline energy use is determined by energy simulation of the Baseline under the operating conditions of the M&V period.	Savings determination for the purposes of a new building Performance Contract, with the local energy code defining the baseline.

Source: (EVO 2007)

When conducting impact analysis of a large population of structures at the whole-building level, and a high level of accuracy is needed, none of these methods alone are sufficient.

Options A and B focus on ECM isolation and therefore are not applicable to whole-building analysis.

Regarding Option C, the protocol states,

> *Option C is suitable only for projects, which do not require a high level of savings accuracy and where there are existing buildings available for comparison which are physically and operationally similar except for the ECMs of the subject*

[6] The California Evaluation Framework, June 2004 and the California Energy Efficiency Evaluation Protocols, April 2006, both extensively reference IPMVP.

building. Even then, the potential for error renders this option suitable for only the most cursory M&V programs.

If a high level of savings accuracy is desired, or "*physically and operationally similar*" structures are not available, Option C cannot be used.

Option D permits accurate evaluation appropriate *for a single structure*. This method could work for a sample, but given the large sample sizes needed for the study, this option was not pursued due to budgetary constraints.

Specific Examples of Evaluation Challenges and Responses

Impact evaluation of the 2004 & 2005 California Statewide Energy Star® New Homes Program (RLW Analytics 2007) used utility billing analysis, equipment sub-metering analysis, and whole building engineering analysis, to assess program energy impacts. The evaluation did not draw upon IPMVP Vol III, Part 1, since the protocol had not been published when the evaluation plan was developed.

Billing Analysis

Generally billing analyses are used in pre/post retrofit scenarios (Brown & Nevius, 2007), which is not the case in new construction since there is no pre condition. The evaluators recognized this challenge, and the results of the billing analysis were not used in the final reported savings estimates, but this approach was attempted to assess its viability for consideration in future evaluations.

A fundamental challenge to this approach was finding a baseline – physically and operationally similar non-participants. In response, RLW took advantage of a previously conducted baseline study (Itron, 2004) and used regression analysis to control for differences between the non-participant (baseline) and participant homes. The analysis considered eleven independent variables including floor area, number of stories, number of occupants, occupancy schedules, and income. While some of this data was available from the Itron baseline study, other information, particularly occupancy, was not, creating a new challenge. A new response was needed, and telephone surveys were developed and implemented to obtain the necessary occupancy and other information. This is an example of a common pattern of evaluation: challenge, response, new challenge, new response, etc.

A practical challenge was the acquisition of billing data. It was determined that some of the participant and non-participant homes received service by two different utilities. As a result it was not possible to obtain all the billing data, and additional responses were developed to overcome this obstacle.

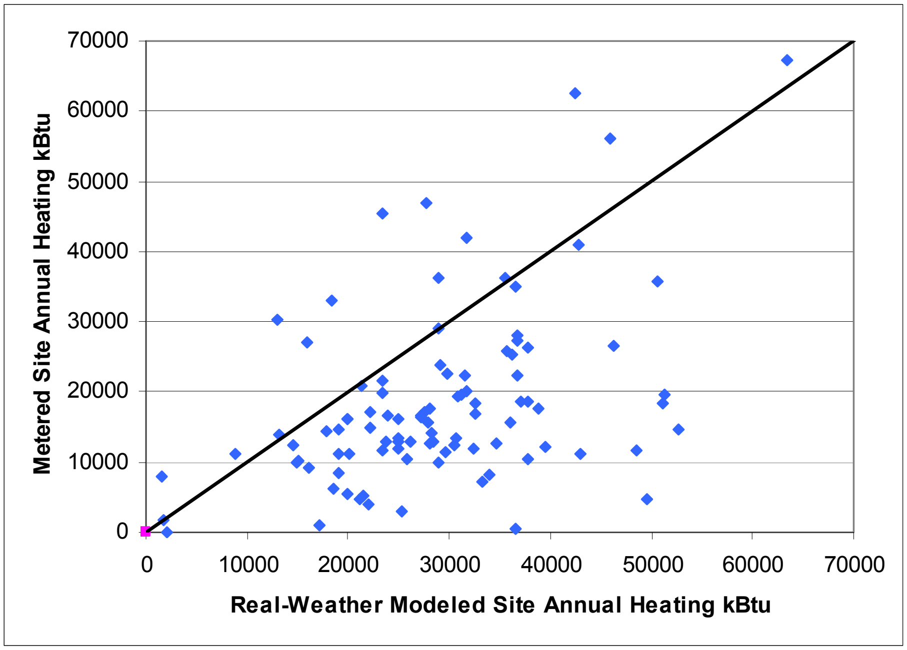
Engineering and Sub-Metering Analysis

RLW Analytics used an innovative approach combining sub-metered participant data to calibrate compliance models. ("Compliance software" is used in California to determine if a new home design complies with the state's building code, Title 24. The software creates an engineering simulation model of each home's design, and compares it to a prescriptive code-compliant baseline home.) Roughly speaking, this is similar to combining IPMVP options C and

D at the project level, coupled with sampling and expansion to yield program level impacts. The results from this innovative combined methodology were used to report final impact results.

To compare modeled to metered data in a meaningful way, the models required adjustment to actual weather and actual as-built characteristics, since actual weather varied from 30-year averages and homes were not always built as planned/modeled. These efforts required two data collection efforts, weather data and on-site field inspection data, and two re-modeling efforts. Combined, these represented significant work and additional practical challenges to achieve the necessary adjustments. Typical results are shown in Figure 3.

Figure 3: Single-Family Metered Annual kBtu Heating Compared to Modeled kBtu Heating



Source: (RLW Analytics, 2007). Notes: Each point represents one home. Modeled energy consumption (x-axis) is weather adjusted and inspection adjusted (models reflect actual as-built characteristics determined by on-site inspections). The reference line shows where modeled energy equals metered energy.

The results show large behavioral variance, consistent with other studies, but by inspection alone it is clear the models are overestimating average heating energy consumption. The study took these results, and similar results for cooling and water heating, and used ratio estimation to calibrate the models. Due to the sample sizes and large metered variance, some meter adjustment calibrations were found to be statistically significant, while others were not.

These are examples of several challenges facing the evaluators. Even after overcoming the challenges of weather adjustments and re-modeling for as-built conditions, the behavioral variance "noise" was still sometimes too great to make meaningful calibrations to the models.

Of the many practical challenges associated with this method, the installation and recovery of hundreds of data-logging meters in homes located throughout California ranks at the top. In addition, the meters had to be left in place for about a full year to capture both heating season and cooling season impacts. This practical consideration alone could eliminate this method during evaluation planning.

Summary

These are just a few examples of the many challenges, both theoretical and practical, of the different implemented evaluation methodologies. The next section summarizes some of these key issues of each method as an example for future evaluation planning purposes.

Summary of Challenges and Considerations for Billing, Sub-Metering, and Engineering Analyses

The following tables summarize each evaluation methodology used in the Energy Star Homes evaluation. Ideally, similar tables would be developed during evaluation planning of other programs to aid in methodology selection. Note that IPMVP does not make a distinction between billing analysis and equipment sub-metering -- they are both considered "metering methodologies" (utility bills are meter based). However, the following summary tables do make this distinction since there are significant implementation and practical differences between these methodologies.

Table 2: Billing Data Analysis

Approach and Advantages	Key Challenges and Considerations	Practical Challenges	Responses to Challenges
• Examine the billed energy use of participants and non-participants (baseline) • Use existing metered (utility) consumption data • Possible to have very large sample sizes (both P and NP)[7] • Possible to acquire several years of data • Evaluation team does not have to wait to "collect" data • Uses "real" data	• Typically used only in retrofit situations • Behavior: large variability of usage (variance can swamp differences) • May be difficult, or impossible, to find physically and operationally similar non-participants • Provides no insight into peak demand reductions unless short-interval metered data are available • De-aggregating billing data (example: cooling) • Behavior biases, positive and negative. (Do occupants of participant buildings behave like those of non-participants?) • Provides little or no insight into how savings occurred • "Real" data does not ensure better results	• Data acquisition: Multiple fuel types provided by multiple utilities • Data quality: utility billing systems sometimes estimate consumption. (Meters are not always read!) • Data interpretation: is it an outlier or a huge/tiny user? • Occupancy changes (homes sold, long vacations, etc.) • Weather variations • Data management – utility billing systems can be complex	• Use regression analysis to statistically control for factors affecting energy consumption (e.g. floor area, occupancy, income, number stories, etc.) • Use billing data de-aggregation methods • Carefully select large samples of participants and non-participants • Careful data QA/QC • Adjust results for typical weather

[7] P, participant; NP. Non-participant

Table 3: End-Use Sub-Metering Analysis

Approach and Advantages	Key Challenges and Considerations	Practical Challenges	Responses to Challenges
• Sub-meter energy usage of specific end-uses • Examine the installed measures in a sample of the participants and non-participants (if NPs are baseline) • Focuses on the affected measures and end-uses to reduce variation and bias from other factors • No de-aggregation efforts needed as in billing analysis • Can provide good insight into peak demand reductions • Uses "real" data	• Large usage variance • Sample size(s) of participants and non-participants, how big for meaningful results? • If used to calibrate models must revise models with actual weather data • Sample bias – are the samples representative? • "Real" data does not ensure better results	• Cost: meters, installation/recovery labor, liability • Time: meters in place for at least a month, season, year, etc. • Timing: installing all meters <u>before</u> period of interest • Data losses from failed equipment and non-recoverable equipment • Change of occupancy (home sold, long vacation, etc.) • Acquiring weather data, if needed	• Sufficiently large samples • Good metering recruitment plan and implementation • Sufficient budget and time • Correct for differences in occupancy and behavior

Table 4: Engineering Whole-Building Simulation Analysis

Approach and Advantages	Key Challenges and Considerations	Practical Challenges	Responses to Challenges
• Use energy modeling to simulate usage and savings • Takes advantage of compliance models[8] already created for code compliance • Can be performed on entire population of participants • Can provide demand reduction information • Can provide insights into <u>how</u> savings occurred	• *Compliance* models not intended to estimate energy consumption • Simulation models can be biased, over or under-estimating energy • Requires existence of baseline homes (representative non-participant homes) • Requires a baseline study • Program participation determined by compliance models, so lacks independent verification	• Trickier for multifamily housing since baseline is more difficult • Homes often not built exactly as modeled, differences found both plus and minus • Data acquisition -- requires obtaining/ modeling representative non-participant homes	• Calibrate models with sub-metered data • Consider different methods for single family and multifamily structures • Conduct inspections to see how construction differs from plans

In the case of engineering models calibrated by metered or billing consumption, the challenges and considerations of the methods should be combined from Table 3and Table 4.

[8] In California "compliance software" is used to determine if a new home design complies with the state's building code, Title 24. The software creates an engineering simulation model of each home's design, and compares it to a prescriptive code-compliant baseline home. The Energy Star Homes evaluation took advantage of these compliance models.

Conclusions & Recommendations

There are many factors to consider when selecting an appropriate DSM-program impact evaluation methodology. No rule-based method exists to select methodologies due to the uniqueness of programs and evaluation goals and constraints. Evaluation professionals must rely on experience and examples of other evaluation efforts to guide methodology selection. In addition to satisfying evaluation goals, evaluators must also assess practical challenges, and consider methods to address these challenges. Guidelines are helpful during the methodology selection process at both the project and program level (EVO 2007), but these can not be relied upon exclusively.

- Early identification of key evaluation challenges aids methodology selection. For example, new construction has no easily identifiable baseline, isolation of ECMs is not possible, and there is likely large variance in occupant behavior.
- Evaluators should be encouraged to re-visit methodology selection at project kick-off meetings, or earlier. (Often awarded through a bid process, evaluators may not have been able to invest sufficient time to vet all evaluation options at the time of proposal writing.)
- Program implementers should not have a hands-off approach during evaluation planning. Implementers have intimate knowledge of the program, available data sources, data quality, and other information that could have a major influence on methodology selection.
- Practical evaluation challenges with possible responses should be considered during the evaluation planning process. Evaluators may find it helpful to create summary tables of each evaluation approach as shown in this paper.
- When estimating savings by all methods, the baseline is half the challenge.
- Behavioral variation must be considered in all projects/programs where human behavior affects energy consumption/savings.
- All metering methodologies are challenged if large behavioral variation exists. Variation of usage of 3x to 6x are common, and as much as 100x have been observed in residential projects. Unless energy savings are expected to be large, evaluators should expect behavioral "noise" to swamp the energy savings "signal."
- Engineering analysis methodologies eliminate the challenges associated with behavioral variance, and can provide meaningful results. However if not calibrated, models can introduce bias, and individual projects are still subject to behavioral variation. "Your mileage may vary."
- Calibration of engineering simulation with metered data is good in concept, but to overcome the noise of behavioral variation may require both large savings and large samples.
- Is there a bias for metered data? While metered data may feel more "real," evaluators must consider what methods will yield the most meaningful results. When options exist, metering is not always superior to engineering analysis.

References

Danny S. Parker, 1990. "Monitored Residential Space Cooling Electricity Consumption in a Hot Humid Climate," Proceedings of the 1990 Summer Study on Energy Efficiency in Buildings, Vol. 9, p. 253, American Council for an Energy Efficient Economy, Washington D.C.

Danny S. Parker, Maria D. Mazzara, John R. Sherwin, 1996. "Monitored Energy Use Patterns In Low-Income Housing In A Hot And Humid Climate." Tenth Symposium on Improving Building Systems in Hot Humid Climates, 316.

Marian V. Brown, Monica J. Nevius, 2008. "Metering the Unmetered Resource, Evaluation Methods for Achieving Diverse Energy-Efficiency Policy Objectives." Consortium for Energy Efficiency

Efficiency Valuation Organization (EVO), 2007. *International Performance Measurement and Verification Protocol, Concepts and Options for Determining Energy and Water Savings*, Volume 1 and 3, San Francisco, CA: EVO.

Brian Dean, Dean Gamble, David Meisegeier, Haider Khan, 2008. "Occupant Energy Index: Understanding How Occupant Behavior Impacts Energy Use." Presented at AESP Conference 2008, ICF International.

RLW Analytics, 2007. *Evaluation, Measurement, and Verification of the 2004 & 2005 California Statewide Energy Star® New Homes Program*. San Francisco, CA: CPUC. http://www.calmac.org/publications/Final_Version_of_04-05_CAESNH_report.pdf

TecMarket Works, 2004. *The California Evaluation Framework*. San Francisco, CA: California Public Utilities Commission

TecMarket Works, 2006. *California Energy Efficiency Evaluation Protocols*. San Francisco, CA: State of California Public Utilities Commission

Itron, 2004. "California Residential New Construction Baseline Study."

Sonderegger, R.C., 1978. "Movers and Stayers: The Resident's Contribution to Variation Across Houses in Energy Consumption for Space Heating," Energy and Buildings, Vol. 1, No. 3, Elsevier Sequoia, Netherlands.