Effective Research Methodology for Smart Grid Enabled Consumer Behavior Impact Assessment

Jordan Michel, Geavista Group
Pam Osterloh, AEP Texas

ABSTRACT

Utilities across the country are assessing the role of smart grid enabled and behavior-based programs in their demand-side management (DSM) portfolios. American Electric Power (AEP) Texas is conducting research that encompasses both smart grid and behavior in the AEP Texas SMART View In-Home Device Research and Development Project (SMART View Project), which is designed to assess the impact of in-home displays on consumer behavior and energy consumption. In this paper, we outline best practices for research design using the AEP Texas SMART View Project as a case study to illustrate critical elements of the research methodology that will allow AEP Texas to identify impacts attributable to the devices. Aspects of the methodology that will be covered include the pre-project pilot, sampling plan, screening methodology, definition of treatment and control groups, and participant surveys. These research practices will be useful to any utility evaluating the role of new technologies in their portfolio.

Background

Like many utilities across the nation, AEP Texas is currently making a significant investment as it upgrades to an advanced metering infrastructure (AMI). Eager to begin quantifying the positive impacts that investment might have on AEP Texas' energy efficiency portfolio, the company decided to allocate a portion of its research and development (R&D) budget to measuring the energy savings impact of in-home displays (IHDs).

Although other utilities have conducted similar studies, their results have varied widely, dependent on the myriad factors that can impact behavior-driven IHD program results (demographics, energy costs, rate plans, climate, IHD capabilities, etc.). Recent meta-analyses published by American Council for an Energy-Efficient Economy (ACEEE) and Electric Power Research Institute (EPRI) found that studies assessing the impact of real-time feedback documented wide-ranging energy savings averages, spanning from -5.5% to 32% (Ehrhardt-Martinez, Donnelly, & Laitner 2010; EPRI 2009). Given the wide range of impacts measured across different studies, this paper focuses not on the results of the AEP Texas research but on the research design that will allow AEP Texas to quantify the impacts of the IHDs they are testing. It is our hope that the research best practices presented here will be of use to other organizations attempting to assess impacts of behavior-based and/or smart grid enabled initiatives, not only for R&D, but for any program that will require impact assessment at its conclusion.

Research Best Practices

Because energy consumption is affected by a multitude of variables and behavior-based savings can be difficult to isolate, it is critical that programs with a goal of assessing behavior impacts be designed with evaluations in mind. The best practices outlined below are common practice in the market research industry; broader adoption of these techniques in the DSM field will lead to improved ability to evaluate the success of energy efficiency initiatives.

Begin with a Soft Launch

A soft launch is a limited release of a program, offer, or recruitment effort that gives the program team the chance to verify assumptions before too many customers are contacted and significant program funds are expended, making it an essential element of any research plan. Although percentages may vary based on program scope and other factors, a typical soft launch may be conducted with a goal of reaching about ten percent of the total projected participants.

For outreach that requires customer follow-up (e.g., surveys, enrollment opportunities), the soft launch allows the program manager to verify response rates and, if applicable, qualifying rates. For example, if a program manager expects to enroll 1,000 customers by sending 20,000 invitations (a 5% combined response and qualifying rate) and discovers after launching the outreach that the combined response and qualifying rate is actually just 2%, the program budget and feasibility could be impacted. Beginning with a soft launch gives program managers the opportunity to confirm assumptions and adjust plans as needed.

The soft launch is especially critical when new technologies are involved. Even when working properly, adoption of new technologies requires customer support. The soft launch provides an opportunity to assess the support needs and plan appropriately before the technology has been distributed to all participants.

A soft launch can also provide an opportunity to catch mistakes. If there is a problem with the study's recruitment survey, it is much better to address the issue when invitations have gone to just a fraction of the study's sample and not the entire mailing list.

Define the Population and Align the Sample to Maximize External Validity

Before designing sampling methodology and screening criteria, it is critical that the population be clearly defined. A statistical population is the group of people to which the results of a statistical analysis can be applied. In some cases, that population may be an entire segment of a utility's customer base (e.g., all of the utility's residential customers), but in many cases it may be more complex (e.g., likely users of a given technology). Defining the population too broadly may dilute the results (i.e., decrease average savings), whereas defining it too narrowly limits the size of the group on which results can be projected.

External validity means that the research results can be extrapolated from the narrow universe of participants to the larger population. The sampling frame is the group of people from which the sample (i.e., participants) is selected. To achieve external validity, the sampling frame must be representative of the population to which it will be extrapolated. If, for example, a study's population is all of a utility's residential customers, the sampling frame should be either the entire population or a representative sample of the population, not a sub-segment that may differ in important ways from the larger population. Alternatively, if the population is defined as a subset of customers that cannot be pre-identified, the program can apply screening criteria to determine which customers align with the study's population. In this situation, however, program results will not be applicable to customers that do not meet the screening criteria.

One important factor in the alignment of the population with the sampling frame is the enrollment process. Programs designed with opt-in enrollment are inherently limited in sampling frame, because participants who choose to participate cannot be assumed to be the same as those who do not. Programs with opt-out enrollment, on the other hand, can have a larger sampling frame and, therefore, can be extrapolated to a broader population.

Include a Control Group to Maximize Internal Validity

Internal validity refers to a researcher's ability to determine a cause-and-effect relationship between an independent variable (e.g., energy efficiency measure) and a dependent variable (e.g., energy consumption). Identifying those relationships, however, can be difficult, especially when the dependent variable can be affected by many different factors. For that reason, it is important to include a control group within the program design. A control group is similar to other program participants in every way possible but does not experience the effect of the independent variable (e.g., no energy efficiency measure is provided).

In order to ensure similarity between treatment and control groups, participants should be randomly assigned. One of the difficulties in establishing a true control has to do with selection bias. In a program with opt-in enrollment, people who choose to participate cannot be assumed to be the same as people who do not choose to participate. Two common research techniques that allow researchers to minimize selection bias are recruit-and-deny and recruit-and-delay. By recruiting members of the control group through the same methodology as other program participants, program managers can ensure that the groups are as similar as possible.

Design Treatment Groups to Isolate Impacts of Research Variables

It is not unusual in energy efficiency programs to combine multiple measures to increase the overall savings impact. However, when designing programs with a goal of assessing savings impacts, it is important to remember that the impact assessment will be more informative if the program is designed in a way that allows for each treatment (i.e., energy efficiency measure) to be assessed in isolation. A matrix structure (see example below) may be a good solution in these cases, because it allows researchers to test both isolated and combined impacts of measures.

Table 1. Matrix Structure Example

		Measure B	
		Present	Not Present
Measure A	Present	A and B	A only
	Not Present	B only	None

In this example, the three groups with a measure present represent various treatment groups, and the group with neither measure becomes the control. Isolating impacts in this way allows program managers to make more informed decisions about program design. If, for example, Measure A in isolation produces 90% of the savings achieved by Measures A and B combined, but costs half as much as the combined measures, the program manager has the data necessary to determine whether the incremental savings justify the incremental cost.

Recruit Sufficient Sample to Achieve Analysis Goals

Sample size refers to the number of participants being evaluated in a study, and it is a critical factor in determining how precise a study's findings can be while remaining statistically significant. However, sample size is also a key cost driver. Utilities must weigh the benefits of increased precision against the costs of recruitment and implementation to determine the appropriate balance. Although sample size calculations are unique to the type of statistical testing being performed, the calculation will likely include the following elements:

- n – sample size; represents the number of participants within each group being tested
- Δ or E – represent the amount of change that must be observed and the tolerable error (+/- value) respectively; these values may be adjusted in accordance with the program's goals
- α and β – represent the probability of error; there is some flexibility in assigning these values within certain generally accepted boundaries
- σ – standard deviation; in the planning stages, this is usually an estimate and is not within the experimenter's control to manipulate (Ott & Longnecker 2010)

Although it is desirable to have a low Δ or E value and low α and β values, lowering these values increases n. Therefore, the program team must carefully assess the goals of the evaluation as well as the feasibility and cost of increased sample to determine the appropriate balance.

Over-Recruit to Account for Dropouts

Once the sample size goals are established, it is important that program managers remember that not everyone recruited in the beginning of the program will end up providing useful data at the program's conclusion. Customers may enroll but not follow through with installation, move before the evaluation period is concluded, or introduce other variables that cannot be controlled for. For these reasons, it is important that programs recruit more participants than required for the final analysis to ensure that sufficient data is available at the study's conclusion. Appropriate rates of over-recruitment vary, but factors to consider include:

- Demands on the customer – How easy is it to participate?
- Length of the evaluation period – How many days/months/years of data do are needed?
- Participant characteristics – Who is being evaluated? Homeowners? Renters? Commercial property owners? Each population will have unique characteristics impacting the likelihood of follow-through.

Plan for an Assessment Period Sufficient for Evaluation of Long-Term Impacts

Allowing time to assess long-term impacts is especially critical for behavior-based savings where persistence cannot be assumed. Savings observed within the first year of participation may be inflated based on short-term engagement, or they may mark the beginning of long-lasting energy savings habits. What "long-term" means for any given program will depend on the utility's goals, but it is important to remember that behavior-based savings cannot be assumed to extend for a period longer than the period during which they were evaluated.

Combine Various Data Sources to Improve Understanding of Impacts

Energy consumption data is, of course, the most essential component in measuring energy savings impacts. However, additional data can greatly expand the opportunities for analysis:

- Weather data can be used to normalize energy consumption, facilitating comparison between groups. Weather normalization, however, may not be necessary if there is a randomly assigned control group.
- Data from other utility programs can be used to identify energy savings not attributable to the program's treatment(s).
- Participant feedback can provide insight into outside factors that may impact energy consumption, as well as a qualitative assessment of the customer experience.

AEP Texas SMART View Project: A Case Study

To illustrate what these best practices look like when applied, a high-level overview of the AEP Texas SMART View Project research methodology is provided below. This research plan illustrates the implementation of the best practices described above as well as providing an opportunity to discuss alternative approaches and opportunities for improvement of the plan.

Internal Device Pilot

In preparation for launch of the full study, a pilot study was conducted with AEP Texas employees to:

- Identify devices that are the best fit for the study
- Assess the level of support that will be required for launch of the project
- Identify common technical issues, so that:
    - preventative measures can be implemented
    - project support staff is prepared to handle common inquiries

Participants in the internal pilot were recruited through an email to local AEP Texas offices. Volunteers received a free device as well as instructions for provisioning the device (i.e., establishing a connection between the device and the meter). Participant support was managed through a dedicated email account, and all issues were logged in a simple tracking system. Several weeks after provisioning their device, pilot participants completed a short survey about their experience installing and interacting with the device.

As a result of the pilot findings, the project team made several key adjustments to the research plan, which will facilitate a smoother rollout to participants. The changes included:

- Device Selection: The pilot revealed that one of the eligible devices had too many barriers to installation and too little visibility to occupants once installed. For these reasons, the project team elected not to include it in the full study.

- Pre-delivery verification: AEP Texas identified two pre-delivery verification strategies that significantly reduced the percentage of participants who encountered technical issues during the device setup process.
- Staggered Delivery Schedule: Based on support needs encountered during the pilot, the project team determined phasing the delivery schedule over a period of 4-6 weeks would keep the support needs of participants at a manageable level.

This internal pilot may be considered a type of soft launch. However, the recruitment phase also began with a soft launch (see below).

Sampling Plan and Screening Methodology

Because IHD programs require an opt-in recruitment methodology, the project team determined that the population for their study would be likely participants in an AEP Texas IHD program. Knowing that AEP Texas could not claim savings for an IHD unless a program participant had chosen to receive one from the utility, the team knew that there was no value in attempting to apply the results more generally.

Participants will be recruited through an online survey to confirm that they meet the project's screening criteria. Because AEP does not have email addresses, targeted postcards with a link to the survey will be mailed to randomly selected single-family residential homeowners with advanced meters. To encourage a high response rate, survey invitations will promote availability of a sweepstakes incentive (e.g., respond to be entered to win one of ten $100 VISA gift cards). Qualifying survey participants will be given an opportunity to indicate their willingness to participate in the AEP Texas SMART View Project. Because knowledge of the survey's purpose could influence the responses provided by respondents, the opportunity to pilot new technology will not be mentioned until eligibility has been confirmed.

In addition to allowing the project to pre-screen participants, recruiting participants through a survey facilitates the recruit-and-deny methodology for establishing a control group. Because qualified participants will indicate their willingness to participate rather than actually applying to be a participant, the project team will be able to confirm that members of the control group are similar to members of the treatment group in every respect other than the treatment.

Recruitment Soft Launch

Because this study relied on a recruitment method never before used by the utility, it was essential to confirm the response rates and qualifying rates estimated in the research plan. The cost and reach of the postcard recruitment effort made the soft launch especially critical; with a plan to send over 100,000 postcards, it was very important that the team knew what to expect and make no mistakes. The postcard outreach was soft launched with 5% of the total anticipated invitations to assess actual response rate and qualifying rate, as well as test multiple recruitment messages.

To ensure the highest possible response rate, the team developed two unique recruitment postcard templates to test during the soft launch. Both templates included essential elements such as the survey url, the sweepstakes offer, and the entity conducting the research. One template was designed to use the least possible copy while the other was formatted as a letter and included additional elements often listed as best practices for survey recruitment, including a privacy disclaimer and an explanation of the research purpose. Each template included a unique identifier in the survey url to track results.

Interestingly, the team found that the more formal invitation was less effective in recruiting homeowners; the response rate was one half of the response rate observed with the more concise invitation. The team also evaluated whether qualifying rates varied based on the postcard received, but no significant difference was observed. As a result, the team adopted the higher response rate template as the basis for the final template.

Although the qualifying rate of survey respondents was quite close to the estimated value of 35%, the response rate for the postcards was significantly lower than projected (0.4% as opposed to 1.5%). As a result, the project team made several modifications to the postcard to increase response rate, including:

- addition of an expected timeframe for response to motivate action
- addition of the utility's logo to increase perceived credibility
- increasing size of the survey url to draw attention and improve readability
- use of colored cardstock (white was used in the initial soft launch) to draw attention

The early discovery of the low response rate also allowed the team to prepare an alternative research design in case the sample size required for the original design could not be achieved. To ensure that sufficient sample could be obtained through this recruitment methodology, a second soft launch was conducted before the postcard recruitment was fully launched (results are pending).

Treatment Groups

Because the project team was interested in determining if the impact of IHDs might be augmented by the distribution of supplemental educational materials, the project uses a matrix structure to define the treatment and control groups. Each treatment is being tested in isolation as well as in combination with other treatments, so that when the project concludes, the team will be able to assess the impact of each individually.

Table 2. AEP Texas SMART View Project Treatment Groups[1]

		Supplemental Communication	
		Present	Not Present
Device Type	Device A	A1	A2
	Device B	B1	B2
	No Device	O1	O2

Participants who meet the screening criteria will be randomly assigned to one of the six groups. The number initially recruited into each group will be 20% higher than the targeted sample size to account for participants who do not successfully set up their IHD or must be removed from the assessment for other reasons.

[1] Please note that this is a preliminary design and is subject to change based on the results of the recruitment effort and availability of eligible participants.

Sample Size

Like many programs (especially programs with opt-in recruitment), the AEP Texas SMART View Project had limitations that required the team to make compromises regarding necessary sample size. Minimum sample size needs for the hypothesis test were calculated based on the following formula for a two-sided test comparing means of independent samples:

$$n = 2\sigma^2 \frac{(z_{\alpha/2} + z_{\beta})^2}{\Delta^2}$$

(Ott & Longnecker 2010)

The α and β values were set at 0.05 and 0.1 respectively. The σ was estimated based on past energy consumption data from the region.[2] The estimation of Δ was based on the savings that would be required to justify the cost of the devices to the utility and the range of savings that could be reasonably anticipated based on previous research. Due to the relatively high Δ/σ ratio, the calculated n was quite low.

Because the calculated minimum sample size was low, a second calculation was used to demonstrate the value of increasing the sample size beyond the minimum required for the hypothesis test. This calculation was based on the confidence interval for comparing means of independent samples:

$$E = z_{\alpha/2}\sigma\sqrt{\frac{1}{n} + \frac{1}{n}}$$

(Ott & Longnecker 2010)

The same α and σ values were used in this calculation. Then, n was manipulated upwards to demonstrate how increased sample size allows for a reduced confidence interval. Weighing the results of this calculation and the interest in precision (as indicated by a reduced confidence interval) against the complications of recruiting participants and the need to control costs, the project team determined that a sample size of 100 participants per cell (plus 20% to account for dropouts) was required and that, when appropriate, cells would be combined during analysis to provide increased precision for comparisons between larger participant groups.

Assessment Period

Due to a hard deadline at the end of 2013, the AEP Texas SMART View Project will include an assessment period of 14 to 16 months. This span of time will allow the project team to assess impacts beyond the first year. It will also include two summers, which is a critical period for managing peak demand in Texas. However, if savings are observed and continue into the final months of the assessment period, the team will not have sufficient data to determine the full persistence of the impacts. (If this occurs, the project team may recommend a follow-up evaluation in subsequent years.)

---

[2] Because the study will use difference-in-differences analysis, σ was based on year-to-year change in electricity consumption, rather than total electricity consumption.

Post-Treatment Survey

After the completion of the assessment period, the project team will conduct a survey of participants (including the control group) to assess:

- Participant perception of IHDs' impact, including:
  - Behavior changes
  - Equipment upgrades and home efficiency improvements
- Participant satisfaction with the device
- Changes in awareness/understanding levels evaluated during initial survey
- Other factors that may have impacted energy usage (renovations, change in number of occupants, etc.)

This survey data will be combined with participation data for any SMART View participants who also participated in other AEP Texas energy efficiency programs. Analyzing this data in combination with the consumption data will allow the project team to develop a deeper understanding of the project's impacts, including:

- Cross-program benefits – Are people with a display more likely to participate in other utility-sponsored programs?[3]
- Qualitative participant experience feedback – Do the participants perceive a benefit?
- Additional factors that impact energy savings – Are certain segments of the population (young/old, high/low income, high/low consumption, etc.) more likely to show savings that others?

Conclusion

There are always constraints that make it difficult for programs to achieve optimal design, and each program team must weigh the tradeoffs. These best practices provide an outline of important considerations for the program design process:

- Begin with a soft launch
- Define the population and align the sample to maximize external validity
- Include a control group to maximize internal validity
- Design treatment groups to isolate impacts of research variables
- Recruit sufficient sample to achieve analysis goals
- Over-recruit to account for dropouts
- Plan for an assessment period sufficient for evaluation of long-term impacts
- Combine various data sources to improve understanding of impacts

While these recommendations are especially important for research and development projects, they can improve the opportunities for assessment of any program that will require verification of energy savings.

---

[3] If such an impact is observed, it will be necessary to discount the savings to avoid double-counting.

References

Ehrhardt-Martinez, K., K. A. Donnelly, & J. A. Laitner. 2010. "Advanced Metering Initiatives and Residential Feedback Programs: A Meta-Review for Household Electricity-Saving Opportunities." Washington, D.C.: American Council for an Energy-Efficient Economy.

EPRI. 2009. *Residential electricity use feedback: A research synthesis and economic framework.* Palo Alto, Calif.: Electrical Power Research Institute.

Ott, R. L. & M. Longnecker. 2010. *An Introduction to Statistical Methods and Data Analysis,* Sixth Edition. Belmont, Calif.: Brooks/Cole.