NON-TRADITIONAL PROGRAM EVALUATION: KEYS TO SUCCESS AND LESSONS LEARNED FROM PG&E'S ENERGY INFORMATION SERVICES MARKET TRIAL

Mary G. Dimit, Pacific Gas & Electric Company, San Francisco, CA
Jeffry J. Erickson, Hagler Bailly Consulting, Inc., Madison, WI

Introduction

This paper will briefly discuss the methodology used to survey participants and estimate the energy impacts in a non-traditional program evaluation at PG&E. This is followed by a discussion of some lessons learned in planning and implementing a non-traditional program evaluation. It will also discuss how the analysis and interpretation of results may be different from more traditional program evaluations. Due to the proprietary nature of this research, no evaluation results will be discussed.

The Energy Information Services (EIS) Market Trial is a joint effort by Pacific Gas and Electric Company (PG&E), TCI, and Microsoft to test a revolutionary in-home energy information system. The first part of this paper concerns Phase Two of the EIS that used a television interface to allow participating residential customers to control their heating and cooling thermostat settings. Participants were also able to schedule lights and other appliances to come on and off at specific times throughout the day, in addition to performing other functions.

This system provided participants with ready access to more information about how their households used energy and allowed them to have a higher level of control over their energy use. The system consisted of small devices that plug in between major household appliances and the electrical outlets, an energy management unit (EMU) at the household meter to record detailed information, and a box that was connected to the television and cable network. This set top box let the user control the connected appliances by using a menu-driven interface on their TV screen. Users could turn their thermostat up or down, turn lights on or off, and schedule their air conditioner and heating system, among other things.

PG&E and Microsoft Corporation are continuing the EIS Market Trial with Phase Three being implemented this summer. Phase Three will try to determine if it is feasible for residential customers to manage household energy use via computers linked to the Internet. The participants will have the same energy management options as in Phase Two with a different interface method. The change to using an Internet interface (via computer and telephone line) instead of the television interface (via set-top box and cable TV) used in Phase Two was due to the slow rate of development of an interactive TV medium.

Methodology

Energy Impact Analysis

Overview. Energy impact evaluations strive to focus on the net effects of a program--the changes in energy use that are attributable to the program. The research design and statistical methods for analyzing energy impacts provide the framework for inferring the net effects from the total (gross) energy impacts. As a result, the research design must include a method to differentiate between program impacts and any naturally occurring changes in energy consumption -- in this case, the level of energy use that would have occurred in the absence of the EIS.

Naturally occurring changes in energy use can either be estimated from a control group of non-participants, or, as was done for this analysis, by assuming that pre-EIS energy consumption patterns for the participants represented the level of consumption that would have occurred post-EIS without the system. Thus, any difference between pre- and post-EIS energy consumption, after adjusting for weather effects, is assumed to be attributable to the EIS. Without a clear trend, any individual estimate of change in consumption is unlikely to be statistically significant, limiting the ability of the analysis model to identify savings in the group as a whole.

This energy impact analysis compared pre- and post-EIS electricity and gas consumption for fifty participants for the period between January 1994 and September 1996. This included a comparison of gas consumption during the winter before and the winter after the system was installed and of electricity consumption during the summer before and after the system was installed.

The availability of weather data and pre- and post-EIS energy consumption data allowed the use of an analysis model known as PRISM to calculate EIS impacts. PRISM, the Princeton Scorekeeping method, is a well-established industry-standard program that calculates weather-normalized energy use by using billing data and long-term temperature data. PRISM produces a weather-adjusted index of consumption called the Normalized Annual Consumption (NAC) for each house and then calculates house-specific changes based on the difference between the NAC before and after the program was implemented. The NAC is an estimate of consumption under average weather conditions. PRISM calculates total program energy impacts by summing the changes from all participants.

PRISM calculates both base-level consumption, which is non-weather-related consumption, and weather-related consumption. PRISM also calculates a reference temperature for each house for the cooling and heating seasons. The reference temperature is the outdoor temperature at which the heating or cooling system turns on. It is generally several degrees lower than the average indoor temperature in the winter and several degrees higher than the average indoor temperature in the summer.

Data Preparation and PRISM Analysis. We used data from three sources for the analysis: PG&E billing records, weather data from April 1983 through October 1996, and participant surveys. PRISM is designed to examine one fuel at a time. We ran PRISM's heating-only model on the gas data and its cooling-only model on the electricity data. Using the EIS installation dates for each program participant, we divided the energy consumption data for each participant into pre- and post-installation periods. In most cases, the pre-EIS period included two complete summer periods and most of two winter periods and the post-EIS period included one summer and one winter period.

PRISM results are particularly sensitive to outliers and estimated readings that are not close to the true readings. As a result, we performed two data cleaning steps using the PRISM software in accordance with PRISM recommendations.

To clean data, PRISM first runs a diagnostic procedure to search for estimated readings. PRISM looks for high-low or low-high pairs of consecutive readings and assumes that the second reading is compensating for errors in the first. It then creates combined readings for the two months. After this, PRISM runs a test of studentized residuals[1] to identify outliers and, if so directed, runs a robust version of its procedures on the outliers. Rather than removing outliers, and thus reducing sample sizes, the robust model assigns weights to each data point with outliers being assigned the smallest weights. In this way, outliers contribute to the analysis but are discounted so they do not overly influence the results.

After adjusting for estimated readings and outliers, we then ran the final PRISM models and examined the results to determine if the data provided reliable results. PRISM suggests using three criteria for eliminating cases from the analysis: an R^2 value of 0.7 or more, a coefficient of variation on the normalized annual consumption [cv(NAC)] of 7 % or less, and a Flatness Index (FI) criteria. The Flatness Index tests for non-weather-dependent consumption. Cases that fail the R^2 and cv(NAC) criteria can be included in the analysis if they pass the Flatness Index criteria.

[1] The studentized residual is the residual divided by an estimate of its standard deviation that varies from point to point, depending on the distance of the value from the mean.

Once PRISM identified the houses that met its criteria, it calculated an impact estimate and reference temperature for each house for the pre- and post-EIS periods. The sum of the impact estimates equals the net program impact.

Surveys

A baseline survey was conducted after all participants had their EIS systems installed. The purpose of the baseline survey was to gather information regarding the participants' current thermostat settings and practices, household equipment and appliances, and attitudes concerning their energy use and related topics.

PG&E's EIS Phase Two final survey was conducted in the fall of 1996. In addition to an update of the baseline information, this final survey gathered energy use data not previously collected from the EIS participants and gave them a final opportunity to offer feedback on the system, including any advice on how to build a successful EIS system. In addition, there were questions regarding different parameters that might affect the participants' electricity and gas consumption. Participants were also asked to be in a focus group in conjunction with this survey.

Data Comparisons

After using PRISM to calculate reference temperatures for each participant, we compared this data to the final survey data to look for correlations or explanations for the energy impacts. We also combined data from the earlier participant baseline survey to examine customer reports of changes in thermostat settings, household characteristics, and EIS system usage.

Keys to Success and Lessons Learned

Following is a discussion of some keys to success and lessons learned in planning and implementing a non-traditional program evaluation and how the analysis and interpretation of results may be different from more traditional program evaluations.[2]

While this section discusses some of the keys to success and lessons learned in the evaluation of new products, many of these issues can also be applicable to new product development, especially of a technology-based product.

Technology-Related Factors

Technological issues. Technical difficulties are normal in the start-up phase of any pilot program. These can range from the product not working at the customer site to the product causing power quality interference with the customer's existing equipment. These difficulties may affect the evaluation results in unexpected ways. For ex-

[2] These lessons were not necessarily learned during nor do they necessarily apply to the PG&E EIS market trial.

ample, customer attrition may be higher than usual in the early stages of a pilot (or any new) program due to customer inconvenience of having to deal several times with the utility to get the product to work properly.

Another important technology-related factor is obtaining good program data. If possible, the more automatically that data collection can occur, the better. In the EIS system, there was an Energy Management Unit (EMU) that automatically collected customer use and other data in real-time. It is also necessary to verify that the data are being recorded properly in the field and can easily be retrieved. Data collection in the field may run into technical difficulties not apparent from the in-house product testing.

Product design. Product characteristics can make or break both customer acceptance and program implementation involving a new product and/or technology. A product must be easily understood and used by a customer and not involve a steep learning curve. That is, the customer must easily be able to learn how to use the new product properly.

The product/technology must be cost-effective. The product must be able to be fielded quickly and cheaply to all participants along with any necessary changes that may occur later. If there is too steep a learning curve on the company side, too much time and money will be spent getting the product implemented. It is not uncommon for a company to develop a product only to find out that the product development costs can never be recouped by product sales. In addition, technology-related products tend to have a short shelf-life due to rapid changes in technology. A product can also be obsolete shortly after it is released due to a competitor's products. Remember eight-track tapes and Betamax videotapes?

And we can't forget those pesky bugs. Many times it is not possible to discover all bugs due to the difference in field versus in-house product testing. However, it is important to avoid implementing/installing a product without sufficient product testing. It is much more expensive to fix a product after it is in production than when it is a prototype model.

Methodological Limitations

Method of analysis. Changes in energy use may not be simply attributed to the difference in participant pre- and post-consumption when additional equipment is involved. For example, in the EIS program, we had to adjust for the electricity consumption of the set top box itself. Of course, impact estimates will be more accurate for those who have gone through a complete heating and/or cooling season after any weather-related system is installed and have ample data prior to installation. Impact estimates are more likely to identify any savings that actually exist when complete pre- and post-program data are available.

Quantitative versus qualitative. Do not misinterpret qualitative data to represent quantitative results. Many pilot programs may not field enough participants to expand any results to the larger population, especially when in the early stages of product or program design. Be careful that others do not misinterpret the evaluation results. For example, don't assume that a decrease in household energy consumption will occur just because the evaluation showed a decrease if the results were not statistically significant due to the small sample size.

Sample design and customer selection. Participants should match the target market of the product/program as much as possible. Customer screening could include such factors as:

- customer or building energy consumption
- demographics (age, education level, income, number of residents)
- technology (customer familiarity or possession of)
- geography
- building characteristics
- "psychological profiles" (such as a customer predisposition toward energy saving or interest in technology)

When planning for the program/product implementation, be sure to consider customer attrition. It is likely to be higher for new products or technologies due to the unexpected, including technical difficulties and customer confusion. When designing the sample, plan for a higher than usual rate of customer attrition. In addition, in any evaluation that requires both pre- and post-consumption data, there will be customer attrition due to causes unrelated to the program or product. These include customers moving, remodeling, changing the number of residents, or use of the building, etc.

To increase customer retention, such items as participant training, a toll-free telephone number, and an instruction booklet may help. Another helpful approach is to begin field testing with "super-friendly" participants. Don't include these participants in test results, as they are not typical participants and will skew any results. There may be the need for several "super-friendly" participants, depending upon how new or different the technology or program is.

Customer responses. Customer response bias can occur for any number of reasons. For example, when asking a participating household to complete a survey, responses may be different from the participant who used the system versus the other household members.

Unexpected factors or considerations can occur no matter how well the evaluation is designed. For example, some participants had the EIS system installed on a secondary TV that they didn't use as often, perhaps because of worries about interfering with their primary television system.

Customer surveys to obtain pre-program (or baseline) information to be compared to post-program actions and/or beliefs are very useful. The timing of a baseline survey is critical. Actions before customers begin participating in a pilot program may be lost by administering the survey too late.

Installation dates. In a pilot program, problems encountered with installation and training are worked out over the course of the pilot. This may result in the "installation" date for some participants (the date used to look for energy savings) not representing the date that the participants could fully use the system.

Customer Feedback

Participant satisfaction. Participants may cite technical problems due to the equipment or program being new. Some participants may find that the instructions are too difficult or confusing and/or the process of using the product to be too cumbersome to operate. On the other hand, other participants may find it interesting to work with the system, just because it is new.

The first installations of a new product or a product using a new technology may have more problems or require more on-site visits or customer interaction than later installs. Naturally, participant satisfaction with the product will be affected. Thus, participant responses from the first installations may need to be weighted differently than later responses when the program is running smoothly.

Customer recommendations. It is important to ask participants for any advice on building a more successful program. This is really useful as they will discover problems and factors that weren't considered in the program development phase. Examples are: how to improve the equipment, increase ease of use, improve system reliability, and eliminate "side effects" such as equipment noise.

The program participants can be the most important source of information about the program or product (provided they are representative of the target market). Use these participating customers to help design the "marketing mix" or the "Four Ps" of the product:

- product differentiation (product characteristics)
- pricing (or rate)
- promotion (marketing communications)
- place (distribution channels)

However, when asking for recommendations from participants, it is important to ask not only what they want but also why they want it. If not, an opportunity may be missed to provide them with a better (more cost-effective, easier to implement, etc.) alternative.