GETTING A GOOD EVALUATION FIT: CUSTOM-TAILORED OR OFF THE RACK?

Bruce Mast and Patrice Ignelzi, Pacific Consulting Services, Albany, CA
Miriam Goldberg, Xenergy, Inc., Madison, WI

Introduction

This paper provides a case study of evaluation strategies applied to two refrigerator recycling programs implemented by Southern California Edison Company (SCE) and San Diego Gas & Electric Company (SDG&E). In comparing evaluation strategies employed for these two programs, the authors explore the implications of adopting standardized methods versus suit-to-fit methods for evaluation quality and cost-effectiveness.

The programs were implemented in comparable customer markets in Southern California and in the same regulatory environment. Evaluation of the SCE program covered the program year 1994, while the SDG&E evaluation was for program year 1995. Regulatory requirements for evaluation of these two programs are spelled out in the California Public Utility Commission's *Protocols and Procedures for the Verification of Costs, Benefits, and Shareholder Earnings from Demand-Side Management Programs*, which specifies, among other things, minimum sampling requirements and documentation requirements. In prescribing acceptable evaluation methodologies, the Protocols heavily favor a comparison of participant and billing data using regression techniques.

Though both programs were developed in the same regulatory environment, they reflected somewhat different goals by their designers and incorporated subtle differences in emphasis. As evaluators, we saw that the differences in their design and operation imposed the need for significantly different impact evaluation strategies. Just as important, we found that the unique nature of refrigerator recycling programs made the standard approaches prescribed by the Protocols less than optimal for evaluating program impacts.

Program Implementation

SCE's Spare Refrigerator Recycling Program

The goal of the Spare Refrigerator Recycling Program included customer satisfaction and market transformation as well as resource acquisition. It was implemented by SCE in the manner of a traditional DSM program. Program promotion was handled in-house and customer leads were developed from SCE's ongoing customer contact efforts. SCE promoted the program in a variety of ways, including regular bill insert, radio advertising, direct mail, and distribution of refrigerator magnets.

SCE contracted with Appliance Recycling Centers of America (ARCA) to handle pick-up and recycling tasks. Customers scheduled appointments to have their working refrigerators or freezers picked up and hauled away at their convenience. Customers participating in the program had the option of receiving either $25 cash or a $50 savings bond. Over 48,000 refrigerators and freezers were recycled through the program during 1994.

SCE contractually committed to providing a minimum number of units for ARCA to pick up. As program implementation began, SCE de-emphasized the importance of screening out customers who wished to replace their old units. SCE determined that rigorous prescreening to prevent these customers from participating would sharply restrict the utility's ability to fulfill its contractual obligations to ARCA. As a result, spare refrigerator removals without replacement accounted for only 54% of the refrigerators collected through the program. About 18% of the collected units were main refrigerators and 33% of the units were replaced.

By relaxing this screen, the program probably achieved a lower net-to-gross ratio (NTG) than it might have otherwise; but it also served as an invaluable complement to SCE's refrigerator incentive program. The incentive program, which provided a financial incentive to customers who purchased efficient refrigerators, generated a number of leads for the refrigerator recycling program. Thus the recycling program assured the full benefit of the incentive program by guaranteeing that the old, inefficient units were taken out of circulation instead of resold or kept in service. The two programs together also promoted better customer relations.

By emphasizing total volume of working units collected for recycling, the program also maximized its effect on the overall supply of used refrigerators being offered for sale in the SCE service territory. To the extent that the program restricted the supply of inefficient refrigerators available on the resale market, it promoted greater energy efficiency in the equipment purchase decisions of people who did not participate directly in the program. In this way the program design aimed to achieve spillover effects with potential long-term market transformation benefits.

SDG&E's Refrigerator Roundup Program

The goal of the Refrigerator Roundup Program was resource acquisition via conservation. It was implemented as a bidding pilot program by Planergy, Inc. Planergy was responsible for all aspects of the program, including pro-

gram promotion, pick-up, refrigerator recycling and disposal, and the verification of load impacts as described in the bidding agreement.

Planergy promoted their program in a variety of ways, which included advertising in local newspapers, distributing program brochures at trade shows and local events, radio advertising, direct mail, the internet, and joint ventures with other companies. SDG&E assisted Planergy by providing leads from the Customer Service Telephone Center and field personnel, and by displaying brochures at company offices and at public exhibits and events.

Planergy fielded inquiries from interested customers and screened them for potential free ridership. Since Planergy's earnings were tied to the number of working refrigerators recycled without replacement, the contractor had every incentive to screen out any customer who wished to recycle a non-working unit or who simply planned to replace an old unit with a newer one. Once an interested customer was determined to qualify for the program, Planergy scheduled a pick-up date at the customer's convenience. Planergy picked up the unit at the customer's home after verifying that the unit was in working condition, that it was not the primary unit in the house, and that it was greater than ten cubic feet in volume. Planergy then issued the customer an incentive payment of $25. Picked-up units were brought back to a central facility, where randomly selected units were metered to determine energy consumption under carefully controlled test-chamber conditions. All recyclable materials were extracted from the collected units and toxic materials were safely disposed of. Through eight months of operation in 1995, the program recycled 2,280 working refrigerators and freezers.

Evaluation Approach

SCE's Spare Refrigerator Recycling Program Evaluation

The evaluation approach for this program was developed to capture both direct and indirect effects of removing used refrigerators from circulation. Refrigerator recycling programs are somewhat unique among DSM programs in that indirect effects have the potential to comprise a significant portion of the total program impacts. This potential stems from the existence of a mature resale market for used refrigerators and freezers, involving transfers via used appliance dealers and direct transfers via garage sales and listings in the classified advertisement section of local newspapers. Thus, an evaluation of net impacts of a refrigerator recycling program must consider not only what the participant would have done with the unit in the absence of the program but also, in the case of avoided transfers, what the recipient would have done to obtain a unit in the absence of the program. An evaluation that ignores avoided transfers will generally understate program impacts, whereas one that takes full credit for avoided transfers will generally overstate program impacts.

Gross savings for the program were based on independent metering study results after reweighting those results to reflect the distribution of equipment age and type recycled under the program. Net savings were derived by applying a net-to-gross ratio to the gross impact estimate. The ratio accounted for the range of alternative actions in the absence of the program, including transfers to other SCE customers; it controlled for equipment that was in use only part-time; and it adjusted for partial unit savings from replacing a unit rather than discarding without replacement, The ratio was estimated using participants' self-reported actual actions and intended actions in the absence of the program. The primary basis for the net-to-gross ratio was a telephone survey conducted for this purpose with 450 program participants.

Developing the net-to-gross ratio from this survey information alone required certain assumptions that were plausible but not verifiable from the primary survey, particularly for attributing savings to the program from avoiding transfers to other SCE customers. To strengthen the analysis with empirical estimates of otherwise subjective factors, the following supplemental data were collected:

- telephone surveys with customers who disposed of a refrigerator outside the program, identified from responses to SCE's appliance saturation survey
- telephone surveys with customers who acquired a used refrigerator, also identified from responses to SCE's appliance saturation survey
- telephone interviews with junk and scrap dealers
- telephone interviews with used refrigerator dealers
- compilation of used refrigerator prices from area newspapers

The supplemental data collection effort provided qualitative information that was important for understanding the impacts of the recycling program on the used refrigerator market. It provided an assessment of the market barriers to disposal of used units outside the program. It shed light on the ultimate fate of units discarded by various means. It clarified assumptions about the geographic scope of the used refrigerator market and how the operation of the recycling program affected the availability of used units for resale. Furthermore, it provided key quantitative information that Xenergy incorporated in the analysis via the following steps:

- Determine the fraction of units that would be kept or discarded in some way. These fractions were determined from the original participant survey.

- For units that would have been discarded in the absence of the program, determine the fractions that would have been destroyed, transferred outside of SCE's service territory, or transferred within the service territory. These fractions were determined from the supplemental survey of discarders.
- For units that would have been transferred within SCE, determine what fraction would have been used as primary units and what fraction as secondary units or spares. Also determine the proportions of alternate actions that took place because the transfer did not occur. These determinations were based on the follow-up survey of customers who acquired used refrigerators in some way.
- For each possible disposition, assign an attribution factor between zero and one, representing the contribution of that outcome to net program impacts.

SDG&E's Refrigerator Roundup Program Evaluation

The contract between SDG&E and Planergy specified in great detail the methodology to be used in evaluating the impacts of this program. This strategy provided several benefits to the utility. First, by tying earnings to performance, it guaranteed that the program would be cost-effective from the utility's perspective. This strategy also minimized potential disagreements between the utility and the contractor over the amount or the basis of the contractor's earnings. And it ensured that evaluation needs would be considered in the program design and implementation. As a result, key evaluation information was collected throughout the implementation phase. Finally, the strategy produced an evaluation report that SDG&E, with only minimal expenditure of resources on its part, could submit to the Public Utilities Commission to document its claim for shareholder incentives for the program.

By contract, free ridership for this program was defined as refrigerators that would have been destroyed or transferred in the absence of the program. In other words, earnings were based only on impacts directly associated with a reduction in the number of operating refrigerators and freezers in participant households. Any program benefit from avoiding the transfer of working units from one household to another was discounted. Because of this contractual definition, the evaluation approach for this program focused exclusively in participant and nonparticipant discard decisions. Gross impacts were determined by extrapolating Planergy's test chamber data from the metered sample to the population of program recycled units and adjusting for the proportion of customers who subsequently acquired a replacement unit. Net impacts were determined by estimating a net-to-gross ratio three different ways:

- using participants' self-reported intended actions in the absence of the program,
- using nonparticipants' actual discard decisions as a proxy for participant discard decisions in the absence of the program, and
- using qualitative choice analysis (QCA) to model the inter-related discard and participation decisions.

The reported NTG ratio was based on the method that produced the intermediate result, in this case QCA. These results were consistent with our ex-ante expectations that the self-report result would be biased downward due to self-report bias and the nonparticipant-discard result would be biased upward due to self-selection bias. These expectations were apparently also shared by the crafters of the evaluation plan specified in the DSM bidding contract.

All three approaches relied on data collected from a telephone survey of 370 program participants and a comparable group of 310 eligible nonparticipants. Identifying eligible nonparticipants who had discarded an eligible refrigerator within the previous year or who could have done so presented a small methodological challenge. To avoid calling thousands of customers, we relied on responses to SDG&E's recent appliance saturation survey. From this, we were able to identify a pool of 1,016 customers who had multiple refrigerators or who had recently discarded units without replacement. However, the appliance saturation survey data were not detailed enough to enable us to determine with certainty whether customers would have actually been eligible to participate in the recycling program. Thus, during the phone interview, nonparticipants were asked a battery of questions to establish more precisely eligibility to participate. In this way 147 of the 457 completed nonparticipant interviews were eliminated for ineligibility. Evaluation results were then developed using sample weights that reflected the relative proportion of SDG&E's residential customer sector that would have been eligible to participate in the program but did not.

Lessons Learned

1. One size does not fit all

Over the years, there has been much debate about the merits of standardized methods versus suit-to-fit methods. The adoption of measurement and evaluation protocols by the California Public Utilities Commission has certainly promoted a degree of standardization among evaluations conducted in that state. Horowitz (1995) argues that increased standardization provides significant benefits to utilities, although he cites no evidence to support his assertion.

The two evaluations showcased here illustrate why, even if standardization was acceptable in the past, it is not likely to be appropriate any more. The two programs discussed are quite similar in focus (removal of second refrigerators) but not in intent (conservation versus market transformation). Had this important difference in intent been ignored, then using the same approach for each program would have resulted in mis-measured results in one or the other. Whichever had been chosen, if appropriate for one program, it would not have been able to measure the real impacts of the other. This is because while a DSM-style program aims to influence customers (demand), a market transformation program aims to influence both customers and vendors (demand and supply). Furthermore, as discussed below, a narrow application of Protocol requirements would have likely resulted in mis-measured results in both programs.

By explicitly recognizing unique program features in the design of the evaluation, we have a better chance of answering the question of a program's impact. Measuring first-year energy impacts is a different evaluation from measuring market effects. To even hope to measure them both requires the use of different approaches. While this lesson is not new, it is more important than ever to clearly define what will count as program impacts before setting out to measure them.

2. A custom-fit evaluation may cost less than using an off-the-rack approach

In considering alternative approaches to evaluating an energy efficiency program, it is often assumed that a customized design will be more expensive than one used for other programs. In evaluating these two refrigerator removal programs, we have found that this assumption is certainly not always (and perhaps not even often) true. A "standardized" evaluation of either of these programs, using the California Protocols methods, for example, would have insisted on a heavy devotion of resources to estimate gross measure impacts and likely use of a billing data, regression-based approach to estimate net program impacts. Such an analysis, to control properly for self-selection bias, would generally entail estimation of a discrete choice model of participation, calculation of an inverse Mills ratio, and then inclusion of this ratio in the billing regression model as a control for systematic differences between participants and nonparticipants. As discussed in Heitfield et al. (1996), the discrete choice model requires virtually the same data collection effort as the Qualitative Choice approach used in the Refrigerator Roundup evaluation. In addition, billing analyses to determine impacts require a large investment of resources required to assemble, validate, and manage the billing data themselves. Thus, a billing analysis would certainly have been more expensive than the adopted approach in the case of SDG&E's program and quite probably would have cost more for SCE's program as well.

Even with the extra cost, it is debatable whether a billing analysis would have produced better results. Keating and Kushler (1995) make a compelling argument that billing analysis may underestimate program impacts from refrigerator recycling programs because of problems that confound the analysis. The authors cite four potential problems:

- Widespread awareness of the purpose of the program … may cause nonparticipants to unplug their second appliances…, which will cause a decrease in consumption in the comparison group. This will appear to a billing analysis to be naturally occurring conservation and result in reduced net savings.
- Participants may unplug their appliances before arranging for their pick-up… This effect will reduce the pre-program participation consumption, reducing the apparent savings in the billing histories…
- The largest potential bias comes from the households who had replaced the primary appliance, but who never previously had a second appliance. Without the program to intervene, they would have become second appliances, even if the only reason would have been…that they were too hard to get rid of. It is logical and desirable for the utility to intervene, but the billing analysis will not represent this dynamic, and the household will only appear to reduce consumption by the difference between the efficiencies of the two appliances—and this can be further muted by the increase in size and services of the new refrigerator.
- A similar effect takes place if the household would not have used the second appliance themselves but would have given or sold the appliance to another customer of the same utility for use as a second appliance.

3. Focus evaluation resources on measuring unknowns with greatest variance

Net program annual savings from recycling a refrigerator or freezer can be decomposed into the following components:

- unit energy consumption in kWh per year
- use intensity prior to recycling, expressed as the fraction of the year the unit was in use
- the probability the unit will be replaced after being recycled

- the probability that, in the absence of the program, the unit would have remained in operation somewhere within the utility service territory

Various metering studies and UEC studies have established the unit energy consumption for a wide variety of refrigerator and freezer makes and models with a relatively high degree of precision. The use intensity, while known with less precision, can be directly observed; that is, it can be measured by asking customers how intensely they actually used the unit prior to recycling. Similarly, the probability of replacement is observable by measuring actual replacement rates. However, the probable operational status of the unit in the absence of the program is highly uncertain and difficult to measure because it cannot be directly observed.

Applying the conceptual approach for strategic evaluation planning described by Pigg, et al. (1995), one would focus the greatest resources on measuring the last component, the probability that the unit would have remained in operation in the absence of the program. This logic formed the basis for the evaluation planning for both projects discussed here. In particular, it led to the decision not to use billing analysis to evaluate these two programs. Billing analysis is a very effective tool for reducing the variance in the measurement of the first three components but it is ill-suited for controlling for differences between actual and hypothetical behavior. The four problems discussed by Keating and Kushler can all be traced to inadequate control for these differences.

As an alternative, the Roundup Program evaluation devoted minimal resources to measuring the first two components of net impacts. Instead, it made use of extensive metering that was part of the program implementation to provide solid estimates of unit energy consumption. It measured replacement via a simple follow-up survey of participants. Evaluation resources were focused on the quantity with greatest uncertainty, the probable operational status of the unit in the absence of the program, expressed as the net-to-gross ratio. For net impacts, an analysis of customers' (participants, nonparticipant discarders, and eligible non-discarders) decision processes and factors was used to determine the likely effects of the program. Using a qualitative choice approach to assess the impacts of the program was both sensible and straightforward. This is because the program impacts were contractually limited to the direct impacts from removing refrigerators or freezers that were not replaced by any other units. This limited the focus to decisions that customers alone made.

For the Spare Refrigerator Recycling Program, where the interest was more about how the program may have transformed the market, we expanded the scope of the inquiry beyond customers' decisions to include an assessment of how the program may have affected the suppliers of refrigerator transfer services. In this case, a qualitative choice approach would have been difficult to apply because the method is not well-suited to modeling the secondary program impacts of market transformation that, by definition, include decisions by a greater number of market players. Again, we devoted minimal resources to measuring unit energy consumption, use intensity, and the probability of replacement. Rather, we focused on the variable with the greatest uncertainty, namely the probability that, in the absence of the program, the unit would have remained in operation somewhere within the utility service territory.

In both cases, we made maximum use of already existing information, which was different for each program, and avoided what would have been more costly billing analyses. More importantly, we were better able to focus on what the evaluation really needed to answer in each case.

4. Leverage existing data and past experience

As noted above, in these two evaluations, we made as much use as possible of information already available: appliance saturation surveys to zero in on eligible nonparticipants, metered data on discarded appliance energy use, terms of services offered by local government recycling agencies, unit energy consumption (UEC) values for refrigerator energy use. This allowed us all to focus on what we really did not know, such as: What is the secondary market for refrigerators like? Is it particularly price- or supply-sensitive? Is there a glut or shortage of units for resale? Were the players in the secondary market greatly affected by or even aware of the utility programs? Where do refrigerators disposed of outside the program usually go? And we could devote more resources to identifying the factors that drive customers' decisions to discard their refrigerators either within or outside the program. We believe that by leveraging all the available information and focusing on what we knew the least about, we were able to provide a more balanced set of results for less cost.

5. Custom-fit evaluation requires custom-fit data collection

Historically, discussions about program evaluation have focused on analytical methods: their effectiveness, their applicability, their feasibility, their cost. While these issues are still relevant, a broad array of well-established methods are available to evaluators today. A fundamental challenge now is what data to collect and how best to obtain them. For both programs discussed, survey instruments were crafted to reflect the particular program design features. When leveraging existing data, special care was required to ensure that the data accurately reflected the programs being evaluated. In conducting the qualitative choice analysis for the Refrigerator Roundup, we were able to use existing data to identify a potential pool of customers for the comparison group but additional screening was required to arrive at an appropriate sample. This extra effort was critical because comparison group definition can have an enormous effect on the results of a qualitative

choice model. Likewise, the Spare Refrigerator Recycling evaluation was able to make use of existing metered data only after reweighting to reflect the actual distribution of equipment recycled under the program.

As utilities shift their attention from resource acquisition to market transformation programs, the set of data we need to collect will increasingly vary from program to program. Appliance recycling involves a different set of players than energy-efficient replacement programs. In the latter, equipment vendors and perhaps manufacturers are key market players along with customers/purchasers. In the former, the market players include junk dealers, charities, government agencies or contractors, and customers as potential recipients as well as discarders of the equipment. This means designing a number of different data collection instruments to reach the right audiences and learn their roles in determining the effect of the program. Knowing where to look for information, how to reach the appropriate sources of information, and even assigning "reliability" factors to the collected data points have increasingly become the focal point of evaluations. As more market transformation programs come on line, this will become even more prominent an issue.

Acknowledgments

The authors are grateful to Athena Besa and Linda Linderman of San Diego Gas & Electric, Jay Zarnikau and Julie Junkin of Planergy, and Marian Brown and Edward Lovelace of Southern California Edison, for the use of their data and information about how the programs were operated. Opinions expressed in this paper are those of the authors and do not represent positions adopted by San Diego Gas & Electric Company or Southern California Edison Company.

References

1. Heitfield, E., B. Mast, P. Ignelzi, and P. Rumsey, "A Comparison of Two Net Analysis Methods Using Data from PG&E's Nonresidential New Construction Program," *Proceedings of the 1996 ACEEE Summary Study on Energy Efficiency in Buildings*, Vol. 6, pp. 6.67–6.72.
2. Horowitz, M. J., "DSM Impact Evaluation Planning: Standardization versus Diversification," *1995 International Energy Program Evaluation Conference Proceedings*, pp. 127–132.
3. Keating, K. M., and M. G. Kushler, "Refrigerator Pick Up Evaluations: Do Billing Analyses Leave You Cold?," *1995 International Energy Program Evaluation Conference Proceedings*, pp. 549–553.
4. Pacific Consulting Services, "1995 Residential Appliance Incentives Program: Refrigerator Recycling First Year Load Impact Evaluation," prepared for San Diego Gas & Electric Company, 1997.
5. Pigg, S., G. Dalhoff, and K. Maini, "Applying Uncertainty Analysis for Strategic Evaluation Planning," *1995 International Energy Program Evaluation Conference Proceedings*, pp. 111–118.
6. Steckel, D., M. L. Goldberg, T. Boertman, M. Kushler, "Throw Out Your Dead! Free Ridership in an Appliance Recycling Program," *1995 International Energy Program Evaluation Conference Proceedings*. pp. 555–560.
7. Xenergy Inc., "Extended Impact Evaluation of the Spare Refrigerator Recycling Program," prepared for Southern California Edison, February 1997.