I can't use a Randomized Controlled Trial – NOW WHAT?

Comparison of Methods for Assessing Impacts from Opt-In Behavioral Programs

Bethany Glinsmann, Navigant Consulting, Madison, WI

Bill Provencher, Navigant Consulting, Madison, WI

ABSTRACT

Randomized Controlled Trials (RCTs) are the gold standard for evaluation; however, many energy efficiency programs are designed as opt-in programs, making a control group impractical or impossible. This paper compares different approaches for estimating the impacts from an opt-in behavioral program. Three model specifications are developed and potential sources of bias from each method are discussed. All three model specifications were applied to help triangulate the estimated impacts and potential for self-selection bias. In this particular application, each model produced relatively similar results, indicating that the results likely do not suffer from self-selection bias. However, this may not always be the case. Future evaluations of opt-in programs that lack a control group should estimate multiple model specifications to better understand the potential for self-selection bias inherent in the program results.

Introduction

Randomized Controlled Trials (RCTs), in which customers are randomly assigned to the treatment and control groups, are the gold standard for evaluation. However, many energy efficiency programs are designed as opt-in programs, making a control group impractical or impossible. Evaluations of measure-based programs, such as weatherization programs, often rely on engineering analysis or deemed savings to quantify program impacts. Behavioral programs do not have this option. Savings result from both measures and behavior change, and the relative influence of each of these savings types is not yet well understood. As a result, evaluations of behavioral programs typically rely on regression analysis using customer billing data.

Programs with a RCT design randomly assign customers to the treatment and control groups, and so the only remaining difference between the two groups is caused by the program itself. Consequently, program impacts are the difference between energy usage for the treatment and control groups and are relatively straightforward to quantify. Use of a RCT design eliminates the potential for self-selection bias. Opt-in programs, on the other hand, often do not have a designated control group, creating the potential for the program impacts to suffer from self-selection bias. Evaluators must use a different set of modeling techniques when evaluating opt-in programs. The remainder of this paper explains self-selection bias and describes three modeling techniques available to evaluators of opt-in behavioral programs.

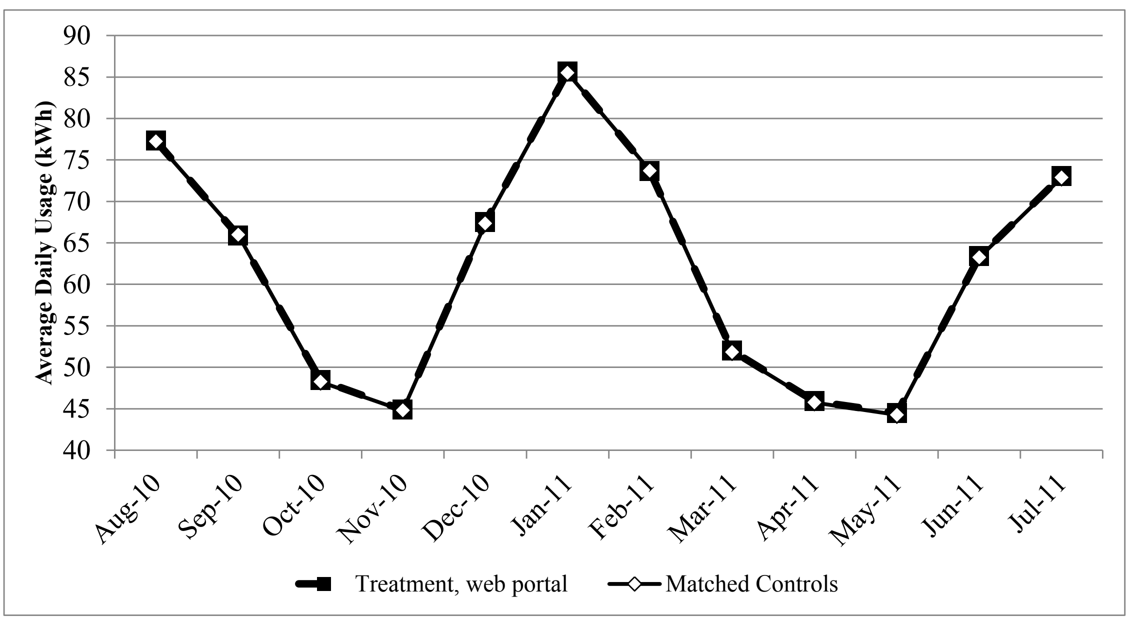
The analysis was conducted on an opt-in portion of a utility behavioral program. The primary portion of the program involved mailing home energy reports to participants. Customers were assigned to the treatment group (those that receive mailed reports) and control group (those that do not receive mailed reports) using a RCT design. This paper focuses on the opt-in portion of the program, in which customers in the treatment group were able to enroll in an online portal to receive detailed information about their energy usage and additional energy saving tips. A total of 947 customers (1.9% of customers receiving home energy reports) enrolled during the first year the portal was available.

The Issue: Self-Selection Bias

Self-selection bias occurs when the decision to enroll in a program is correlated with the customer's energy usage. Suppose, for example, that customers that are interested in energy efficiency are more likely to enroll in an energy efficiency program. Given their pre-existing interest in energy efficiency, these customers likely would have taken energy-saving actions even in the absence of the program. That is, their decision to enroll in the program is correlated with a change in their energy usage. Savings directly attributable to the program must be disentangled from savings that would have occurred in the absence of the program – a difficult task for program evaluators.

Model A: Matched Controls

Model A relies on the use of matched controls to estimate program impacts. For each participant a matched control household was selected from a large pool of non-participants. Selection was based on the minimum sum of squared differences (SSD) in monthly energy usage[1] for the twelve bills prior to the program period. The non-participant with the minimum SSD was selected as the matched control for each participant.[2] Differences in mean monthly usage between the participants and matched controls were statistically tested. None of the monthly differences were statistically significantly different than zero at the 90% confidence level, indicating that the average monthly usage patterns for the participants and matched controls are similar before the program began. Any differences during the program period are attributable to the program itself or to self-selection bias. Figure 1 shows the average usage for both participants and matched controls during the pre-program period. Note that the two lines are virtually indistinguishable.



Source: Navigant analysis

Figure 1: Average Daily Usage During Pre-Program Period, Participants and Matched Controls

Once the matched controls are selected, the next step is to estimate the regression model. The regression model predicts average daily usage during the post-matching period (the period beginning with the receipt of the first home energy report) as a function of participation in the program and average daily

[1] The matching process used average daily energy usage for each bill cycle, where average daily energy usage is equal to the total energy usage in the bill cycle divided by the number of days in the bill cycle.
[2] Matches were selected with replacement, so that a non-participant could be selected as the match for multiple participants.

usage during the pre-program period. Only customers that enrolled in the web portal and their matched controls are included in the regression. Formally, the regression equation is:

Equation 1: Model A

$$kWh_{it} = \alpha + \beta * PreUsage_{it} + \gamma * Participant_i + \delta * Post_{it} + \chi * Post_{it} * Participant_i + \epsilon_{it}$$

Where $\alpha, \beta, \gamma, \delta, \chi$ are parameters to be estimated, *i* indexes the customer, *t* indexes the monthly bill period, and

kWh	= The average daily usage for customer *i* in bill period *t*.
PreUsage	= The average daily usage for customer *i* during the pre-program bill cycle in the same month as bill period *t*.
Participant	= A binary variable indicating if customer *i* receives home energy reports.
Post	= For web portal participants, a binary variable indicating if customer *i* has enrolled in the web portal in bill cycle *t*. For matched controls, this variable indicates if their matched participant has enrolled in the web portal in bill cycle *t*.
ϵ	= The model error term.

Recall that the program has two components: a paper report (sent to a randomly selected group of customers) and an opt-in web portal. The regression model must account for savings from both components. Programs with only an opt-in component require a simpler model, in which the last two terms of the regression equation are omitted.

The first term of the model, PreUsage, is included to reflect the set of variables used in the matching process. The modeler could choose from a number of variables selected to reflect the pre-program usage, such as the annual usage in the pre-program period. Here the variable PreUsage captures the average usage in the corresponding month from the pre-program period. For example, for July of the program period, PreUsage is equal to the usage in July of the pre-program year.

The γ parameter captures the impact of the home energy reports. Because treatment and matched control customers' usage matched prior to the program, we can attribute any difference in usage after the program begins (and prior to enrollment in the web portal) to the home energy reports. In this application, the estimate of program savings from the home energy reports (γ) is consistent with (not statistically significantly different than) the results of a separate analysis focused on quantifying savings from the home energy reports.

The δ parameter captures the average effect *among matched control customers* of being in the period in which their matched participant has enrolled in the web portal, the "post web portal period." For example, if a customer enrolled in the web portal in response to a shock (say a high bill), the δ parameter captures the change in energy usage resulting from the matched control customer's response to this same shock (say purchase of an energy efficient appliance). In this particular application, δ was positive but not statistically significant. The sum $\delta + \chi$ captures the average effect *among participants* of being in the web portal period. The direct effect of the web portal on energy usage is captured by the χ parameter. In other words, this parameter captures the *difference-in-difference* in average daily energy usage between the participants enrolled in the web portal and their matched controls across the pre- and post- web portal period.

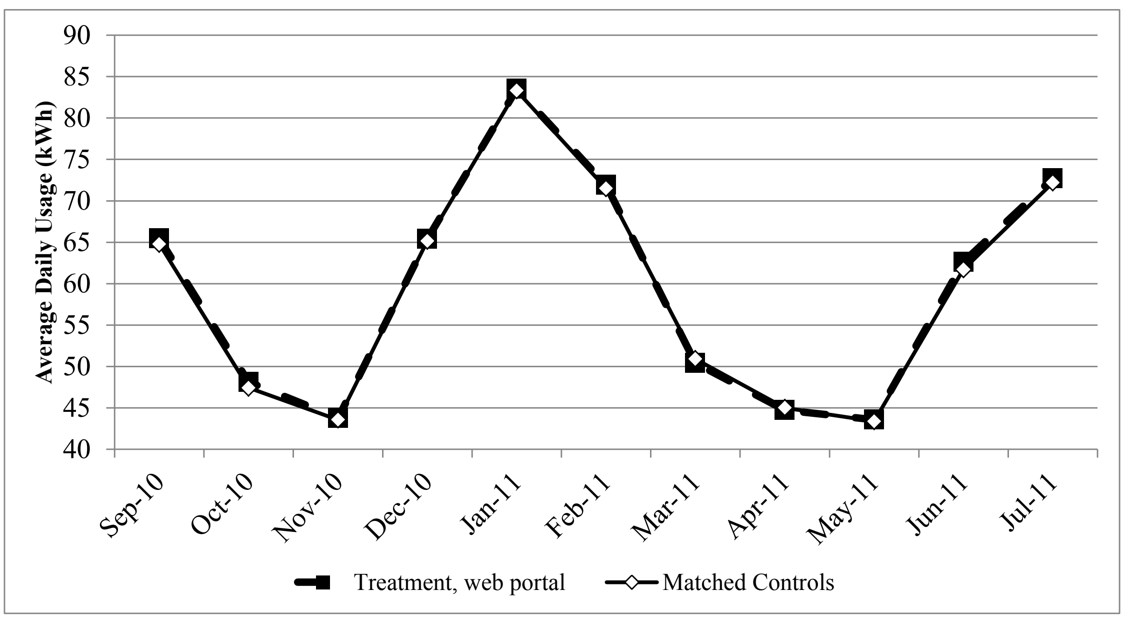
Despite matching on energy usage patterns prior to the program, the potential for self-selection bias still exists. The type of customer that enrolled in the web portal could respond differently to exogenous factors that occur during the program period.

Model B: Matched Controls Selected from Late Participants

Model B is similar to Model A in that it relies on the use of matched controls to estimate program impacts. However, in this model participants are split into two groups based on the date they enrolled in the web portal. Whereas matched controls were selected from a pool of non-participants for Model A, for Model B matched controls for early participants are selected from the pool of late participants. Selecting matched controls from a pool of customers that enrolled in the same program at a later date reduces the potential for self-selection bias.

Recall that self-selection bias occurs when the decision to enroll in a program is correlated with the customer's energy usage. By selecting matched controls from a group of customers that eventually enroll in the program we have eliminated the potential for self-selection bias based on unobserved characteristics that do not change over time but are correlated with energy usage (such as usage level, demographics, or building type). However, the decision of *when* to enroll in the program might also be correlated with the customer's energy usage, and so self-selection bias is still a potential problem. For example, customers with electric heat might enroll in the web portal during the heating season, when their electric bills are the highest.

The late participant with the minimum SSD was selected as the matched control for each early participant.[3] Differences in mean monthly usage between the participants and matched controls were statistically tested. None of the monthly differences were statistically significantly different than zero at the 90% confidence level, indicating the monthly usage patterns for the participants and matched controls are similar before the program began. Any differences during the program period are attributable to the program itself or to self-selection bias. Figure 2 shows the average usage during the pre-program period for both early participants and matched controls selected from late participants. Note that the two lines are virtually indistinguishable.



Source: Navigant analysis

Figure 2: Average Daily Usage During Pre-Program Period, Early Participants and Matched Controls Selected from Late Participants

[3] Matches were selected with replacement, so that a late participant could be selected as the match for multiple early participants.

Once the matched controls are selected, the next step is to estimate the regression model. Similar to Model A, the regression model predicts average daily usage as a function of participation in the program and average daily usage during the pre-program period. Formally, the regression equation is:

Equation 2: Model B

$$kWh_{it} = \alpha + \beta * PreUsage_{it} + \gamma * Participant_i + \epsilon_{it}$$

Where α, β, γ are parameters to be estimated, *i* indexes the customer, *t* indexes the monthly bill period, and

kWh	= The average daily usage for customer *i* in bill period *t*.
PreUsage	= The average daily usage for customer *i* during the pre-program bill cycle in the same month as bill period *t*.
Participant	= A binary variable indicating if customer *i* is an early web participant.
ϵ	= The model error term.

Note that this model includes fewer variables. We are comparing participants currently enrolled in the web portal to those that eventually enroll in the web portal; both groups also receive the home energy reports. Any difference in usage is attributable to participation in the web portal and is captured by the γ parameter.

This model is only applicable to the first part of the program period; that is, the time period during which the early participants enroll in the web portal. Consequently, estimation of the program impact relies entirely on the impact for the early participants (the first half to enroll) during the first part of the program period. If impacts differ for early and late participants, by time in the program, or by season then the model estimate of savings will be biased.

Model C: Variation in Adoption (VIA)

Similar to Model B, Model C does not rely on energy usage data for non-participants. Developed by Harding and Hsiaw (2012), the Variation in Adoption (VIA) model exploits the fact that participants enroll in the web portal at different times of the year. A set of binary variables indicating the calendar month accounts for monthly variation in energy usage exogenous to the program, such as weather. Program impacts are allowed to vary by time in the program via a set of binary variables indicating the number of months since program enrollment. This set of variables allows impacts to ramp up during the first few months of program participation, a common occurrence for behavioral programs. Finally, the VIA model accounts for trends in energy consumption prior to program enrollment with a set of binary variables indicating months until program enrollment. Formally, the regression equation is:

Equation 3: Model C

$$kWh_{it} = \alpha_i + \sum_{m=1}^{M} \delta_m * Month_{mt} + \sum_{p=0}^{12} \beta_p * ProgramMonth_{pit} + \sum_{r=1}^{22} \gamma_r PreProgramMonth_{rit} + \epsilon_{it}$$

Where δ, β, γ are parameters to be estimated, *i* indexes customer, *t* indexes the monthly bill period, and

α_i	= A customer-specific fixed effect (constant term).
kWh	= The average daily usage for customer *i* in bill period *t*.
Month	= A set of M binary variables indicating the calendar month, taking a value of 1 if bill period *t* falls in calendar month *m*.
ProgramMonth	= A set of 13 binary variables indicating the number of months since program enrollment.
PreProgramMonth	= A set of 22 binary variables indicating the number of months until program enrollment.
ϵ	= The model error term.

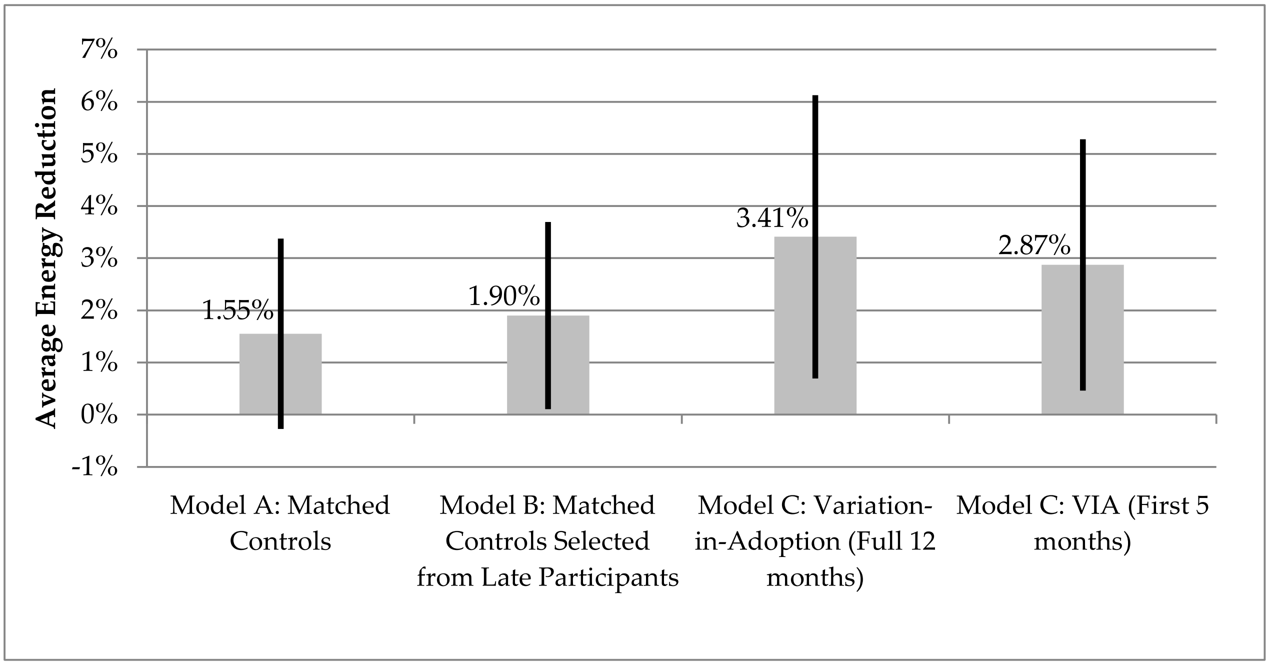
The β parameters capture the program impacts. The set of program month variables allows program impacts to vary by time since enrollment in the program, but does not allow for seasonal variation in program savings. The model assumes that savings in the 5th month of the program are the same whether that month occurs in January or July. The VIA model could be expanded to include a set of interaction terms between program participation and calendar month or season, if the data allow. Note that the month variables capture the effect of the printed reports, since all customers started receiving reports during the same month.

The γ parameters serve as a built-in test of the maintained assumption of the model: early and late participants are similar with respect to energy consumption patterns prior to enrollment in the web portal. If the maintained assumptions are valid, we do not expect that savings before activation will be statistically different than zero, because the customer has not yet enrolled in the program.[4]

Results

Estimation of the three models discussed above resulted in similar estimates of program savings. Savings range from 1.55% to 3.41%, but these differences are not statistically significantly different at the 90% confidence level, as shown in Figure 3. The matching methods (models A and B) resulted in the lowest estimate of savings, while the VIA model (model C) resulted in the highest estimate of savings. These differences are at least partially related to the structure of the models, as discussed below.

[4] Note that using a 90% confidence level implies that random chance could cause one out of ten parameters to be statistically significantly different than zero.

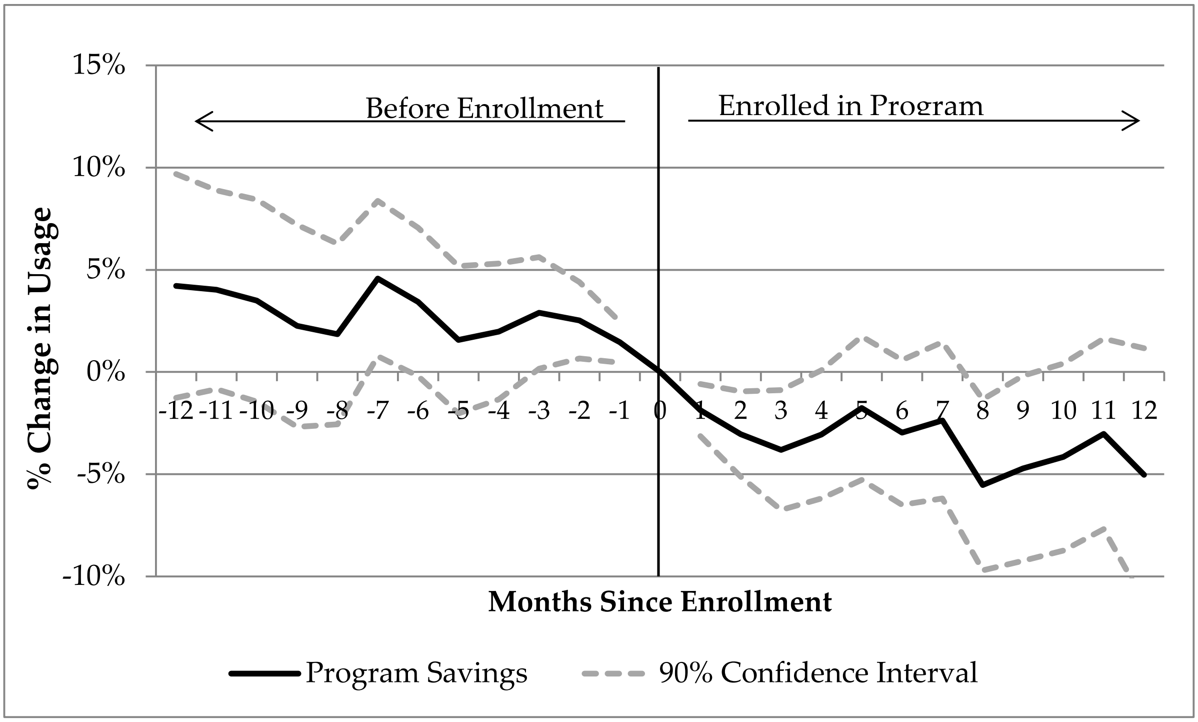


Source: Navigant analysis

Figure 3: Average Savings Estimates with 90% Confidence Intervals

For the models relying on matching methods, each observation has the same weight. Because customers enroll in the program at different times, there are more observations for the earlier months of program enrollment than the later months. For example, all customers are in the program for at least one month, so the number of observations corresponding to usage during the first month of the program is equal to the number of participants. However, the number of observations for each program month decreases as the number of months since enrollment increases. Only customers that enrolled during the initial month that the web portal was available have observations for a full year post-enrollment in the web portal. If savings ramp up during the first few months of program participation, as is typical for behavioral programs, then the estimate of savings does not represent annual savings for a participant *in the program for an entire year*. Instead, it represents the average savings for participants *during the first year of the program*, some of which have been in the program the entire year and others have been in the program for only one month. Essentially, savings during the first months of the program are over-weighted, while savings during the last months of the program are under-weighted in the calculation of average savings. This may appear to be a subtle distinction, but can have a large impact on the estimate of savings.

Alternatively, the VIA model explicitly estimates how program savings change with the number of months since enrollment. The fact that there are more observations for the first program month than the last program month affects the standard error (a measure of uncertainty) on program savings; the more observations, the smaller the standard error. Summing the program month parameters gives an estimate of annual savings for a participant in the program for an entire year. Savings during each month of the program are equally weighted, although savings from the later months of the program have greater uncertainty. Figure 4 displays the monthly savings estimates and 90% confidence interval resulting from the VIA model (model C). Monthly usage is compared to usage during the month of program enrollment (month 0). Note that usage was higher prior to enrolling in the web portal and decreased after enrolling in the web portal, indicating program savings occur. Furthermore, savings increase with the number of months in the program. However, three of the twelve parameters corresponding to months before program enrollment are statistically significantly different than zero. This indicates that the maintained assumptions of the model could be violated, and results from this model should be interpreted with caution. Figure 3 displays the savings estimates from Model C for the full twelve months of the program (3.41%) as well as the savings estimates for the first five months of the program (2.87%). The latter is more comparable to the results from Models A and B.



Source: Navigant analysis

Figure 4: VIA Model Estimate of Program Savings with 90% Confidence Interval

Table 1 summarizes the strengths and weaknesses of the three models discussed in this paper.

Table 1: Comparison of Models

Model	Strengths	Weaknesses
A: Matched Controls	• Includes all participants • Includes the entire program period • Simple regression model	• Selection bias could occur if participants and non-participants respond differently to exogenous factors • Must obtain billing data for a large pool of non-participants • Impact estimate is weighted towards the first months after enrollment
B: Matched Controls Selected from Late Participants	• Reduced potential for selection bias because matched controls eventually enroll in the program • Simple regression model • Only requires billing data for participants	• Program impact is estimated using only the first half of participants and the first half of the program period • Impact estimate is weighted towards the first months after enrollment
C: Variation in Adoption	• Includes all participants • Includes the entire program period • Only requires billing data for participants • Model specification captures variation in program savings by time since enrollment • Impact estimate equally weights all months during the program period • Implicitly tests maintained assumptions of model	• More complex regression model • Basic model specification does not account for seasonal variation in savings

Conclusions and Recommendations

Self-selection bias occurs when the decision to enroll in a program is correlated with the customer's energy usage. To mitigate the potential for this bias, program evaluators should apply alternative model specifications, three of which were discussed in this paper: selection of matched controls from a pool of non-participants (Model A), selection of matched controls from participants that eventually enter the program (Model B), and Variation in Adoption (Model C). Models A and C incorporate data for all participants and the entire program period, whereas Model B includes only the first half of participants and the first half of the program period. Models A and B produce an impact estimate that is weighted towards the first months after enrollment, while the impact estimate resulting from Model C equally weights all months during the program period. Model A could suffer from selection bias if participants and non-participants respond differently to exogenous factors that occur during the program period. The risk of selection bias is reduced in Models B and C because the models do not rely on data for non-participants. However, there is evidence indicating that the maintained assumptions of the model could be violated, and results from this model should be interpreted with caution. In this particular application, each model produced results that were not statistically different from each other, indicating that the results likely do not suffer from self-selection bias. However, this may not always be the case. Future evaluations of opt-in programs that lack a control group should estimate multiple model specifications to better understand the potential for self-selection bias inherent in the program results.

References

Harding, M., and A. Hsiaw. 2012. Goal Setting and Energy Efficiency. Paper presented at the AEA 2013 Annual Meeting, San Diego, Ca., January 4-6.