Energy Savings Over a 3-year Opt-in Rewards-based Residential Behavioral Program

Carly McClure, University of WI-Madison and Navigant Consulting Inc, Madison, WI
Bill Provencher PhD, University of WI-Madison and Navigant Consulting Inc, Madison, WI

ABSTRACT

This study evaluates savings for a residential behavioral program at Commonwealth Edison (ComEd), a large Midwestern US utility, implemented by C3 Energy and the Illinois Citizens United Board (CUB).[1] The program is a web-based, opt-in program designed to generate energy savings by providing customers with information about their energy usage, tips to reduce energy consumption, and reward points for energy savings. We compare three quasi-experimental methods to estimate savings, with the understanding that finding similar savings from the three methods confers "convergent validity" on the estimates. The first method is the variation-in-adoption (VIA) approach used by Harding and Hsiaw (2013), in which program savings are estimated using only data from program enrollees, with late enrollees serving as controls for early enrollees. The second and third methods are matching methods that draw on the same set of program enrollees and their 1:1 non-program matches, but the two are distinguished by the method used to estimate savings. The first is regression with pre-program matching (RPPM) described in Ho et al. (2007) and the second is matching with bias correction (MBC) introduced by Abadie and Imbens (2011). For both of these, matching is based on Euclidean distance in monthly energy use over a 12-month pre-program period. A 2-month pre-program "test window" comparing the average use of program customers and their matches provides a proxy test for selection bias, which is always a concern with opt-in programs. The three methods generate similar estimates for program savings: 3.81%, 3.86%, and 3.57% for the VIA, RPPM, and MCB approaches respectively.

Introduction

In recent years, many US utilities have begun using behavioral programs to achieve energy conservation goals and requirements (ACEEE, 2013). The State and Local Energy Efficiency Action Network (SLEEAN) put out a report in 2012 detailing methods for evaluating behavioral programs which has become a practical guide for evaluation across the US. The report ranks several methods based on their internal and external validity, suggesting that the choice between different models is clear-cut and can be done a-priori with little knowledge of the program involved (Figure 1).

Star Rating	Condition
★★★★★	**Randomized Controlled Trial** results in unbiased estimates of savings.
★★★★☆	**Regression Discontinuity** results in estimates of savings that are likely to be unbiased if done correctly.
★★★½☆	**Variation in Adoption with a Test of Assumptions** could result in biased estimates of savings.[60]
★★★☆☆	**Propensity Score Matching** could result in biased estimates of savings.[61]
★★☆☆☆ Not Advisable	**Non-Propensity Score Matching** could result in biased estimates of savings.
★☆☆☆☆ Not Advisable	**Pre-Post Comparison** could result in very biased estimates of savings.

Source: SLEEAN, 2012

Figure 1. Star ratings ranking different methods of evaluation for energy conservation programs

[1] Acknowledgements: The authors would like to thank Commonwealth Edison for allowing us to use the data from this program in this paper.

Randomized Controlled Trials (RCTs) are broadly recognized as the "gold standard" for program evaluation (behavioral or otherwise). However, many utilities do not implement RCTs for their behavioral programs, in particular when a program requires customers to opt-in in some way, such as by signing up on a website or installing a device in their home. In such a case, some interested customers would need to be turned away from the program to make the RCT possible (e.g. to form the control group) and utilities are loath to upset these customers. Some programs are getting around this barrier by using a recruit and delay strategy in which some interested customers are not allowed to enroll for the evaluation period, which is an RCT, but can enroll later on. However, often utilities are not conducting RCTs, but rather are expecting program evaluators to estimate savings from opt-in programs ex-post using quasi-experimental techniques.

The problem for evaluation of programs which lack random assignment is selection bias, the idea that customers who choose to enroll are different from those who do not. In energy conservation programs, for example, it is possible that people who enroll in the program are more concerned about saving energy, or money, than people who do not enroll. Hence, it is possible that, to some extent, the program participants would have saved energy regardless of the program. As suggested by Figure 1, there are several quasi-experimental methods available to evaluate non-RCT programs. This paper empirically compares three quasi-experimental evaluation methods – the variation-in-adoption (VIA) method and two different matching methods – for the C3-CUB Energy Saver program implemented by C3 Energy and the Illinois Citizens United Board (CUB) for the utility Commonwealth Edison (ComEd).

Often the choice of evaluation method for an opt-in program is dictated by the available data or the set-up of the program. For example, the VIA method requires a program with rolling enrollment, while matching requires usage data from a sufficiently similar pseudo-control group who never enrolls in the program. Fortunately, for the C3-CUB Energy Saver program we have the data to use either of these evaluation methods. SLEEAN (2012) recommends using the VIA approach, rather than matching, without any further exploration of the program. However, in this paper we show that our three evaluation methods estimate very similar program savings – 3.81%, 3.86%, and 3.57% for the VIA and two matching methods respectively – and we offer pseudo-tests of selection bias for each model. Our results suggest that the choice between models for the evaluation of opt-in behavioral programs is not as clear-cut as suggested by the SLEEAN report; rather than picking a model based solely on SLEEAN's rankings, we recommend that program evaluators carefully consider the models available to them and make the decision between them based on the pseudo-tests of selection bias we present. Better yet, evaluators can run several structurally different models to confer "convergent validity" on their results.

Description of the Program and the Data

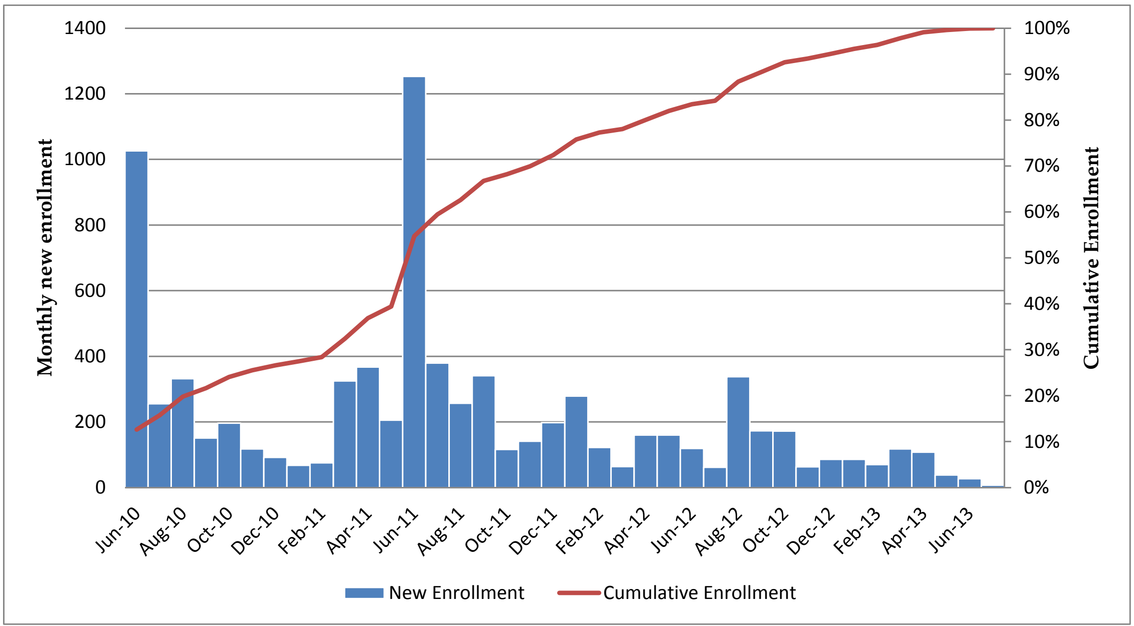
The C3-CUB Energy Saver program is a web-based, opt-in program designed to generate energy savings by providing customers with information about their energy usage, tips on how to reduce energy consumption, an opportunity to set energy saving goals, and reward points for saving energy[2] that can be redeemed at local retailers. The information on savings, tips, and rewards points can be accessed via the program's web portal. The first time they visit the web portal, customers fill out a short survey about their home so that the suggested tips and estimated yearly savings from following those tips can be tailored accordingly. Through the web portal customers can also access figures showing how much energy they have saved and they can spend their rewards points.[3] For sample screenshots from the web portal see Harding and McNamara (2011).

[2] Participants earn 2 points for each kWh of energy they save.

[3] Examples of items that can be purchased using reward points include a $10 gift card for a flower retailer (100 points) and a 15% off coupon for an online store (75 points). However, the rewards and the number of points needed to buy them vary through time and by retailer.

In addition to having access to the web portal, customers receive monthly emails indicating whether or not they saved energy that month compared to their usage in the same month of the previous year. The emails also include the customer's to-date savings since joining the program, their rewards point balance, and two energy saving tips. The tips are seasonally focused, for example a May email would include summer related tips such as turning down the A/C. For sample screenshots of the monthly emails see Harding and McNamara (2011).

The C3-CUB Energy Saver program began in June 2010 and by July 2013 (the last month for which we have data), 8,809 people[4] had at some point been enrolled in the opt-in, web-based portion of the program. Figure 2 presents monthly enrollment and cumulative percentage enrollment since the program's inception. Enrollment surged at the start of the program in June 2010 and at the program's 1-year anniversary in June 2011; both of these events were well-publicized and a concerted effort was made to enroll households in these months.[5]



Source: Navigant analysis

Figure 2. Monthly enrollment and cumulative percentage enrollment, June 2010-July 2013

We received monthly billing data for all program participants and control customers for the period of September 2008 to July 2013. We calculated average daily usage (in kWh) by dividing the monthly usage by the number of days in the bill. We also received the date of enrollment in the program for each participant. Accounts were removed from our analysis for several reasons. We removed accounts which had enrollment dates prior to June 2010 (which were flagged as test accounts to ensure the portal was properly functioning by C3 Energy) and those missing an enrollment date altogether. Additionally, we removed observations of usage that were missing a bill year or had a bill year prior to 2008, observations that were outliers,[6] and observations that had irregular bill lengths.[7] This left us with valid data for 8,138 participants.

The VIA analysis used data from all 8,138 participants. For the matching methods we also removed accounts which had less than 12 months of data in the matching period to improve the quality of our matches. This left us with 6,973 participants in the matching analysis. We also

[4] The program was widely marketed, but it is difficult to definitively state a participation rate, except to say that ComEd has approximately 2.5 million residential customers.

[5] The marketing activities in these months were focused on driving customers to the web portal to enroll in the program not on generating energy savings in and of themselves.

[6] We defined outliers as observations with average daily usage more than one order of magnitude from the median usage in the targeted sample for the analysis. This removed approximately 1% of the data.

[7] We defined irregular bill lengths as bills with less than 20 or more than 40 days.

received data for over 160,000 potential controls that were drawn from ComEd's customer base. We matched on Euclidean distance over a 12-month pre-program matching period with replacement (described in detail in the next section) and selected 6,551 unique control customers as matches. This resulted in 6,973 matched pairs in the matching analysis. Table 1 shows summary statistics of pre-program average daily usage in kWh for the participants in the VIA analysis, the subset used in the matching analyses, the entire pool of potential controls, and the selected controls.

Table 1. Table of Summary Statistics

	Participants in VIA Analysis	Participants in Matching Analyses	All Potential Controls	Selected Controls
Mean Avg Daily Usage in kWh (Pre-program)	24.83	25.73	36.90	26.22
Median Avg Daily Usage in kWh (Pre-program)	20.47	20.92	30.44	21.52
Standard Deviation of Avg Daily Usage in kWh (Pre-program)	18.49	19.53	29.37	19.31
Number of Accounts	8,138	6,973	160,537	6,551

Source: Navigant analysis

Table 1 clearly shows that the pool of all potential controls had higher average daily usage than the participants (36.90 kWh versus 24.83 and 25.73 kWh). This in itself is not a matter of selection bias, since it pertains to observable variables that can be used to condition the dependent variable using matching, regression analysis, or both. The group of potential controls also has higher variance in usage than the participants (a standard deviation of 29.37 versus 20.47 and 20.92), which occurs both because the group of potential controls is larger and because the participants are clearly a subset of the total population who have low usage. Once we match, the selected controls are much more similar to the participants (average daily usage of 26.22 kWh), which reduces the likelihood that selection bias is biasing our estimates of program savings. In the next section we present some pseudo-tests for selection bias in each of the quasi-experimental methods we use.

Description of the Evaluation Methods

The three methods presented below have been selected for comparison because, though they have slightly different data requirements, they are applicable to a broad range of opt-in energy conservation programs. The VIA method exploits the variation in the timing of sign-up in programs with rolling enrollment, i.e. programs where not everyone starts at the same time, and late enrollees provide the counterfactual outcome for early enrollees. In matching, the counterfactual is provided by a group of individuals who never enroll in the program but who are found to be sufficiently similar so as to act as controls for the enrollees. The matching methods are distinguished by the statistical method used to estimate savings. The first is regression with pre-program matching (RPPM) described in Ho et al. (2007). The second is matching with bias correction (MBC) introduced by Abadie and Imbens (2011).

Variation-in-Adoption (VIA)

This method takes advantage of the differential timing of program enrollment by customers to identify program savings. It essentially takes the perspective that the best comparison group for customers enrolled at time *t* is those that enroll later in the program period. The method uses only program participants to estimate savings, with late enrollees essentially serving as controls for early enrollees. It relies on the assumption that, controlling for both customer and monthly fixed effects, neither energy use in month *t*, nor energy savings *s* months into the program, are correlated with the timing of program entry.

The method uses a fairly simple, but flexible, linear fixed effects regression model of energy consumption by households. The base model casts monthly electricity consumption as a function of a household-specific fixed effect, month/year fixed effects, and the time-distance from activation (both pre-activation and post-activation). This is a two-way fixed effects model that indirectly accounts for all time-invariant customer characteristics, and all month/year factors affecting all customers (such as weather and the inflation rate). Formally we have,

Model 1[8]

$$ADU_{kt} = \alpha_k + \beta_t + \sum_{j=-\bar{m}}^{\bar{m}} \gamma_j D_{kt}^j + \varepsilon_{kt}$$

where,

k: household index
t: calendar month index
j: months before or after enrollment date index

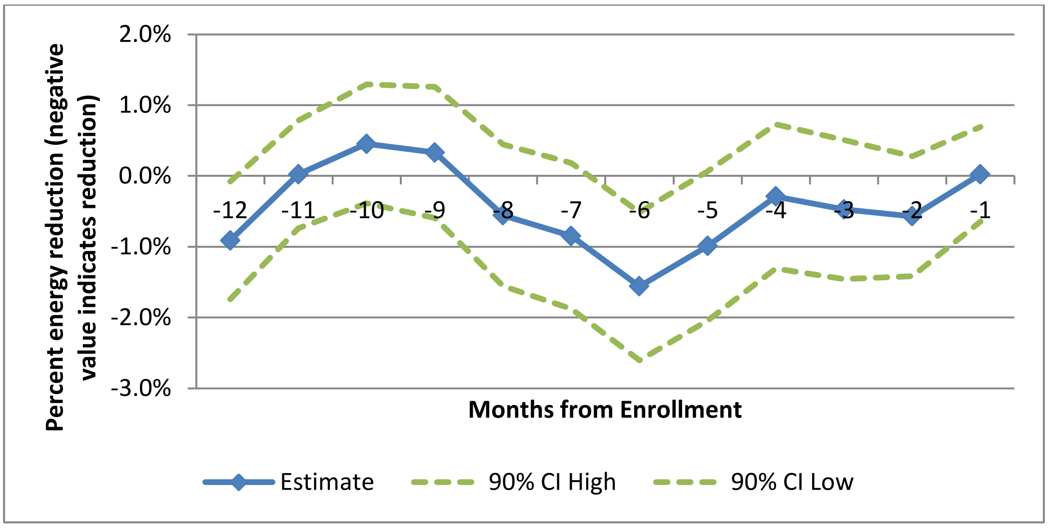
ADU_{kt}	=	Average daily energy use by household *k* in month *t*;
α_k	=	Household-specific constant (fixed effect);
β_t	=	Month/year specific constant (fixed effect);
D_{kt}^j	=	A 0/1 indicator variable that takes a value of 1 if month *t* is the *j*[th] month before/after household *k* activates the web portal. Month $\bar{m}=0$ is the month before enrollment;
γ_j	=	Coefficient on the indicator variable D_{kt}^j;
ε_{kt}	=	The cluster-robust error term for customer *k* during billing cycle *t*. Cluster-robust errors account for heteroskedasticity and autocorrelation at the customer level.[9]

The underlying assumption of the VIA approach is that, after controlling for customer fixed effects, customers who are *j* periods from enrollment are the same on average as customers *j+s* periods from enrollment, where *s* can be negative or positive. So, for instance, customers who are 4 months from enrolling in January 2011 are the same on average as customers who enrolled 6 months prior to January 2011. An important feature of the model is that it reveals, via the values of γ_j for $j<0$, whether customers are likely to start reducing their energy consumption as they approach enrollment,[10] after controlling for monthly fixed effects (an indication of selection bias). If they are not reducing their energy use, and customers are the same on average regardless of enrollment date (again, after controlling for customer fixed effects), then $\gamma_j=0$ for all $j<0$.

Making use of this quasi-test for selection bias, Figure 3 graphs the "average program savings"[11], γ_j, on a percentage basis in the months before program enrollment. In eight months the program effect is negative, in two months the effect is positive, and in two months it is virtually zero. In only one month (T-6, indicating 6 months before enrollment) is the program effect statistically different from zero at a 90% confidence level. At this confidence level, chance alone would cause an

[8] For a formal discussion of the assumptions of this model, please refer to section 1 of the appendix.
[9] Ordinary Least Squares (OLS) regression models assume the data are homoskedastic and not autocorrelated. If either of these assumptions is violated, the resulting standard errors of the parameter estimates are likely underestimated. A random variable is heteroskedastic when the variance is not constant. A random variable is autocorrelated when the error term in one period is correlated with the error terms in at least some of the previous periods.
[10] For example, a customer might have an exceptionally high bill and decide to start saving energy before enrolling.
[11] These are the "program savings" estimated by the model in the pre-program period. Therefore, we expect that these estimates are zero and estimates that are different from zero are not actually savings from the program but rather would indicate that the model is inappropriate for evaluating this program.

average of one month out of ten to be statistically significant. We conclude that the results are reasonably consistent with the assumption that there is no significant program effect before the start of the program, and thus no significant evidence of selection bias.



Source: Navigant analysis

Figure 3. Estimated average percent reduction in energy use due to the program in the 12 months before program enrollment[12] (negative values indicate energy savings)

Matching

The matching methods rely upon a set of matched comparison households to estimate program savings. The pool of non-participant households available for matching consisted of 160,537 ComEd residential customers. In program evaluation, the basic logic of matching is to balance the participant and non-participant samples by matching on the exogenous covariates known to have a high correlation with the outcome variable. Doing so increases the efficiency of the estimate and reduces the potential for model specification bias. Formally, the argument is that if the outcome variable Y is independently distributed conditional on $\boldsymbol{X}$ and D, where $\boldsymbol{X}$ is a set of exogenous variables and D is the program variable, then the analyst can gain some power in the estimate of savings and reduce potential specification bias by assuring that the distribution of $\boldsymbol{X}$ is the same for treatment and control observations. In this evaluation, the outcome variable is monthly post-program energy use, and the $\boldsymbol{X}$ with the greatest correlation to this is energy use in the same month of the pre-program period, $PREkWh_{kt}$, where k indexes the customer and t indexes the month.[13] For further discussion of matching on $PREkWh_{kt}$ refer to section 2 of the appendix.

This section describes how the matches are formed and the two different methods for evaluation after the matching, RPPM and MBC, are described in the next two subsections. The two methods draw on the same set of matches for the comparison group, but differ in their use of a structural model to estimate program savings. The MBC approach is less parametric, using regression analysis to correct for bias in differences between participants and their matched comparisons. The RPPM method, by contrast, treats matching as a "pre-processing" stage of the analysis and assumes that monthly energy use in the post-program period can be modeled as a linear regression function involving participants and their matches.

---

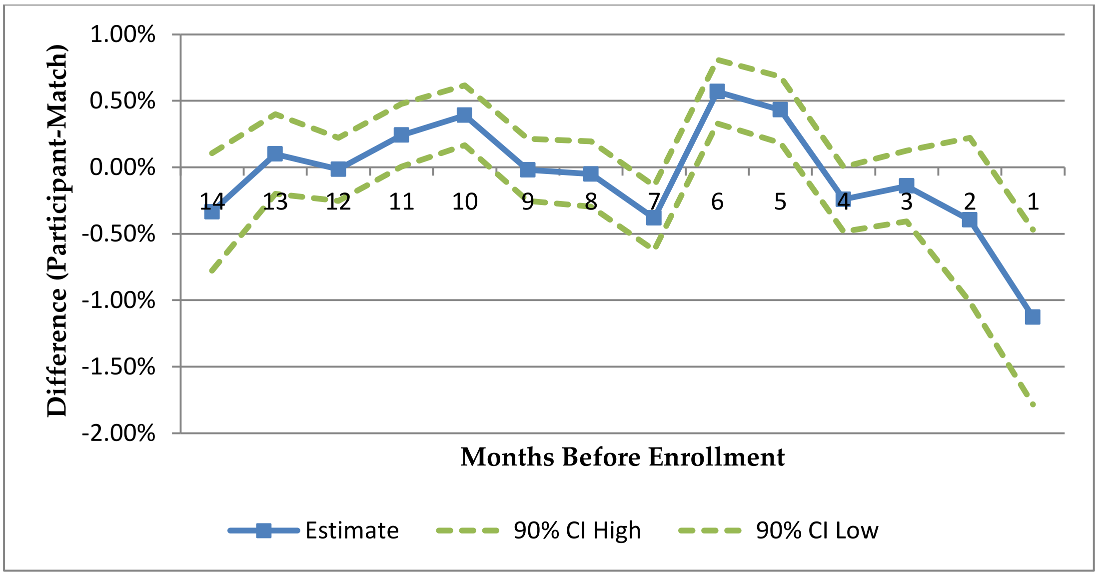
[12] In the VIA model months prior to -12 are estimated along with the 12th month before enrollment.

[13] Correlation between past and present energy usage in our sample is 0.85. We do not match on other demographic variables; such as income, age of household head, or family size; for several reasons. First, we do not have this data available. Second, with such a high correlation between past and present energy usage these other variables are unlikely to be useful in predicting energy usage when past energy usage is included. Finally, this is only an issue if demographic differences are observed on average across participants and their matches.

For each program participant with monthly billing data extending at least 14 months before program enrollment, energy consumption in each month in the period spanning 3-14 months before program enrollment (a twelve month period) was compared to that of all customers in the available control pool with billing data over the same 12 months. For the sake of expositional clarity below, we denote by t_k=0 the month t in which customer k enrolled in the program, with t_k -1 denoting the month before enrollment, t_k +1 denoting the month after enrollment, and so on.[14]

The basis of the comparison is the difference in monthly energy use between a participant and a potential match, D_{PM} (**D**ifference between **P**articipant and potential **M**atch). The quality of a match is denoted by the Euclidean distance to the participant over the 12 values of monthly D_{PM} used for matching; that is, denoting by SSD the sum of squared D_{PM} over the matching period, the quality of the match is denoted by $SSD^{1/2}$. The non-participant customer with the shortest Euclidean distance to a participant was chosen as the matched comparison for the participant. Matching was done with replacement, and this process resulted in 6,973 participants and 6,551 unique matched controls.

It is not possible to statistically test for selection bias, but Imbens and Wooldridge (2009) present a test that is suggestive (hereafter called the "IW test"). In the current context, the logic of the test is that in the absence of selection bias there should be no difference between participants and matches in average energy use outside of the matching period and outside of the program period. A simple implementation of the test is to determine whether, given matching based on months t_k -3 to t_k -14, average D_{PM} in months t_k -1 and t_k -2 is practically or statistically different than zero. Figure 4 presents the average difference in average energy use of participants and their matches over the pre-program period in percentage terms, with 90% confidence intervals. The figure illustrates two important points. First, on average the energy use by matches is very similar to that of program participants. Second, the mean difference in energy use is not statistically different from zero in t-2 (90% confidence level), but is statistically different from zero in t-1, leaving ambiguous the issue of selection bias in the sample. In other words, in period t-1 there is statistical evidence that participants used less energy than their matches, which could be due to discrepancies in the program start date for some participants, but also raises the possibility that on average participants were more inclined than their matches to reduce energy as they entered the program (that is, energy savers were self-selecting into the program), in which case the estimate of program savings would be biased upwards.



Source: Navigant analysis

Figure 4. Average difference in monthly energy use before program enrollment in the matching (14-3) and IW test (2-1) periods, participants and their matches, with 90% confidence intervals

[14] Customers with missing bills during the designated matching period [t_k -14, t_k -3], but whose billing data extended past 14 months before program enrollment, were matched based on their most recent 12 bills before t_k -2 (that is, starting three months before enrollment and working backwards in time).

After selecting the matches we used two methods to evaluate program savings, RPPM and MBC, which can both be interpreted as using regression analysis to control for any remaining imbalance in the matching on $PREkWh_{kt}$. If, for instance, after matching the participants use slightly more energy on average in the pre-program period than their matches then, for both the RPPM and the MBC approaches, including $PREkWh_{kt}$ as an explanatory variable in a regression model predicting monthly energy use during the post-program period keeps this slight difference in pre-program energy use from being attributed to the program.

Regression with Pre-program Matching (RPPM). In the RPPM approach the development of a matched comparison group is viewed as a useful "pre-processing" step in a regression analysis to assure that the distributions of the covariates (i.e., the explanatory variables on which the output variable depends) for the treatment group are the same as those for the comparison group that provides the baseline measure of the output variable. This minimizes the possibility of model specification bias. The regression model is applied only to the post-treatment period, and the matching focuses on those variables expected to have the greatest impact on the output variable.

We matched participant and control customers on energy use during the pre-treatment period, as described above, and then estimated the following model for all post-program observations:

Model 2[15]

$$ADU_{kt} = \alpha_{0t} + \alpha_1 PREkWh_{kt} + \alpha_2 Treatment_k + \varepsilon_{kt}$$

where:

ADU_{kt}	=	Average daily energy use by household *k* in month *t*;
α_{0t}	=	Month/year specific constant (fixed effect);
$Treatment_k$	=	A 0/1 indicator variable, taking a value of 1 if customer *k* is a participant, and 0 otherwise;
$PREkWh_{kt}$	=	average daily electricity use by household *k*'s match during the pre-program month that is the same as month *t*. For instance, if household *k* enrolled in August 2011, $PREkWh_{kt}$ for June 2012 is June 2011;
ε_{kt}	=	The cluster-robust error term for customer *k* during billing cycle *t*. Cluster-robust errors account for heteroskedasticity and autocorrelation at the customer level (see footnote 9).

In this model α_2 indicates average daily savings from the program. We include a monthly fixed effect to account for unobserved time-related factors, such as weather, that affect all customers.

Matching with Bias Correction (MBC). The second matching method follows the approach summarized in Imbens and Wooldridge (2009) and applied in Abadie and Imbens (2011). In this model, the effect of the program in month *t* is the difference between the energy use of participant *k* and its estimated counterfactual (baseline) consumption. The estimated counterfactual consumption is the average consumption of its matched household amended to reflect differences between participants and their matches in the covariates ***X*** affecting energy use. Formally we have,

Model 3[15]

$$Savings_{kt} = ADU_{kt} - ADU_{kt}^{C}$$
$$ADU_{kt}^{C} = ADU_{kt}^{M} + \hat{\beta}\left(PREkWh_{kt} - PREkWh_{kt}^{M}\right)$$

[15] For a formal discussion of the assumptions of this model, please refer to section 1 of the appendix.

where:

ADU_{kt}	=	the average daily electricity use by household *k* during month *t*;
ADU_{kt}^{C}	=	the estimated counterfactual energy use by household *k* during month *t*;
ADU_{kt}^{M}	=	the energy use by household *k*'s match during month *t*;
$PREkWh_{kt}$	=	average daily electricity use by household *k* during the pre-program month that is the same as month *t*.
$PREkWh_{kt}^{M}$	=	average daily electricity use by household *k*'s match during the pre-program month that is the same as month *t*.
$\hat{\beta}$	=	the factors used to adjust household *k*'s energy use for differences between household *k* and its match in the value of $PREkWh_{kt}$.

The values of the adjustment factors $\hat{\beta}$ used in Model 3 are derived from a regression model applied to the post-program period, estimated using *only* the matched comparison households. In the current analysis the regression model used for adjustment purposes is identical to Model 2 except that the variable *Treatment* is dropped, as the model is applied only to the matched comparison households. Formally,

$$ADU_{kt} = \beta_{0t} + \beta_1 PREkWh_{kt} + \varepsilon_{kt} \text{ ,}$$

To apply this regression equation to Model 3, we define $\hat{\beta} = \hat{\beta}_1$. We estimate this regression separately for each month of the program year, generating twelve values of $\hat{\beta}_1$.

Results

The results across the three methods are very similar (Table 2). The MBC model gives the lowest but most precise percent savings estimate of 3.57% (Standard Error (SE) = 0.21%). The VIA and RPPM approaches give very similar percent savings estimates of 3.81% and 3.86% (SE = 0.59% and 0.42%). The details of these results are presented in the next three subsections.

Harding and Hsiaw (2013) studied the first year and a half of this program (June 2010-December 2011) and estimated average annual savings of 4.4% using the VIA approach. Harding and Hsiaw also broke out the savings by the goals customers set and found that customers setting realistic goals (defined as a 0-15% reduction in energy usage) saved nearly 11%, however consumers did not actually set a reduction goal as a percentage of their energy use but simply picked actions they wished to undertake and Harding and Hsiaw converted this into percentage savings. Based on their sample, Harding and Hsiaw also concluded that savings might diminish dramatically after 6 months. By contrast, our analysis, which is based on many more program customers, a longer program time horizon, and uses three different methods to evaluate savings finds slightly smaller program savings and indicates no loss of savings over the first 3 years of the program.

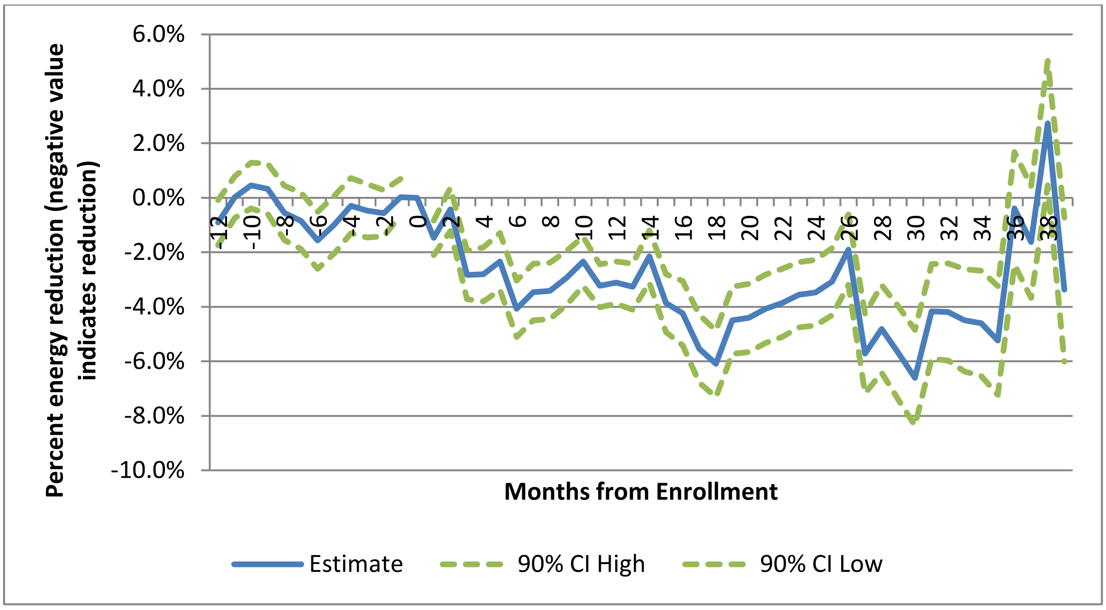
Table 2. Summary of Results

Type of Statistic	Method VIA	 RPPM	 MBC
	(standard errors in italics)		
Number of Participants used in analysis	8,138	6,973	6,973
Average Percent Savings	3.81% *0.59%*	3.86% *0.42%*	3.57% *0.21%*
Average kWh savings per customer per day	0.985 *0.152*	1.037 *0.117*	0.956 *0.056*

Source: Navigant analysis

VIA

Figure 5 plots the average daily savings by month from the VIA model for pre- and post-enrollment months. There is a substantial drop in energy use after the program begins, and this drop appears to deepen as customers enter the second year of the program. There is some evidence that the effect of the program weakens moving into the third year, but this must be interpreted with caution, for two reasons. First, the increase is an amplification of a 12-month cycle of reduced energy savings that suggests the model assumption that differences among customers are not correlated with the timing of enrollment is somewhat questionable. Second, the amplification is likely associated with the fact that relatively few customers (1,613 who enrolled prior to August 2010) have been in the program for more than three years.[16]



Source: Navigant analysis

Figure 5. Estimated average percent reduction in energy use due to the program (negative values indicate savings)

We calculated savings over the entire post-enrollment period by taking a weighted average of the monthly treatment effects by the number of customers enrolled in each month. The average percent savings per enrolled customer is 3.81% (SE = 0.59%). This is an average daily savings of 0.985 kWh per customer (SE = 0.152).

RPPM

The RPPM approach used for this evaluation does not estimate savings by month; rather it just results in one estimate of savings for the entire post-program period. Parameter estimates for the two variables of interest in Model 2, $PREkWh_{kt}$ and $Treatment_k$, are presented in Table 4.

Table 4. Parameter Estimates for RPPM Model (Model 2)

Parameter	Coefficient	Standard Error	t-statistic
PREkWh	0.76842	0.00806	95.28
Treatment	-1.03656	0.11739	-8.83

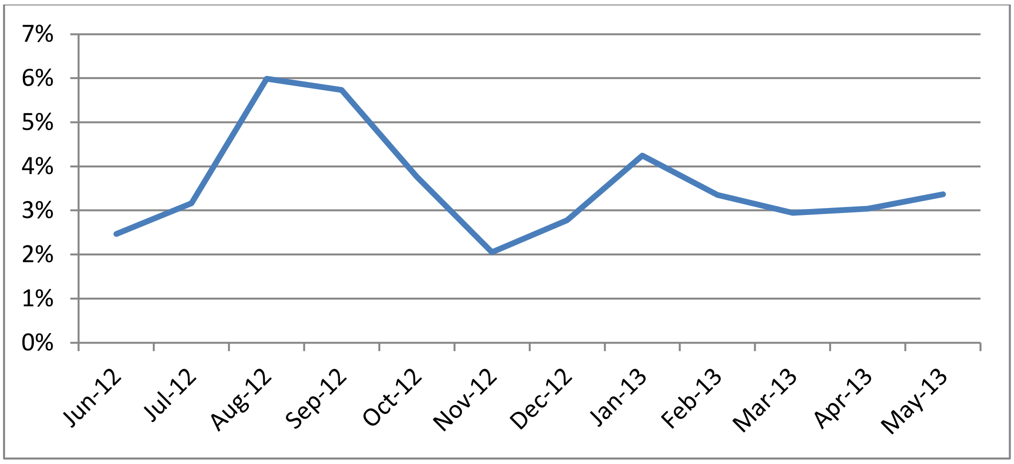
Source: Navigant analysis

[16] Detailed coefficient estimates for this model are presented in the Navigant analysis which is available online at: http://ilsagfiles.org/SAG_files/Evaluation_Documents/ComEd/ComEd%20EPY5%20Evaluation%20Reports/ComEd_C3_EMV_Report_PY5_2014-01-21%20Final.pdf

The parameter estimate on Treatment of -1.037 is the average reduction in daily electricity usage in kWh per customer (SE = 0.117). This leads to an average percent saving of 3.86% per customer (SE = 0.42%).

MBC

The average daily savings by month using the MBC approach are presented in Figure 6. On a percentage basis, estimated savings are highest in August 2012 and peak again in January 2013. It deserves emphasis that the month is the bill month, with August bills averaging as many days in July (the latter half of July) as days in August (the first half of August), and September bills averaging as many days in August as days in September.[17]



Source: Navigant analysis

Figure 6. Average percent monthly energy savings, MBC model

Similar to the VIA approach, we find an average savings estimate from the MBC model by taking a weighted average of the monthly savings estimates by participants enrolled in each month. This results in an average daily savings estimate of 0.956 kWh per customer (SE = 0.056) and an average percent savings of 3.57% (SE = 0.21%).

Conclusion

For the web-based behavioral electricity conservation program we evaluate we find very similar savings estimates across three different models which confers “convergent validity” on our results from a statistical point of view. We find some evidence of selection bias in the matching estimate, as seen by our pseudo-test for selection in the two months between the matching period and the start of the program. The test of the VIA approach assumption in which we examined the monthly treatment effects in the pre-program period to see if they were statistically different from zero showed no evidence of violation of the assumption that late enrollees are good controls for early enrollees. Therefore, our preferred specification is the VIA model in which the program savings estimate is 3.81% (SE = 0.59%). To further explore the selection bias we see in the matching model, we would need to obtain something like a brief questionnaire to discern selection bias which could be developed and administered to new enrollees upon enrollment. Without a process orientation involving a survey we are unable to evaluate what actions customers are taking to save energy or which household characteristics are associated with the most savings. With the data we had, we can

[17] Detailed coefficient estimates for this model are presented in the Navigant analysis which is available online at: http://ilsagfiles.org/SAG_files/Evaluation_Documents/ComEd/ComEd%20EPY5%20Evaluation%20Reports/ComEd_C3_EMV_Report_PY5_2014-01-21%20Final.pdf

only evaluate the average treatment effect, i.e. the average per customer energy savings. We are looking into obtaining such a questionnaire or survey for future years of the program.

Savings of over 3% are quite high for behavioral programs similar to this one, such as Home Energy Report programs, which provide customers with extra information about their monthly energy usage (ACEEE 2013). Additionally, while most Home Energy Report programs generate higher savings for households with high usage, this program primarily has enrollees with lower than average usage; thus it is able to reach a portion of the population underserved by other behavioral programs. ComEd was pleased with the results of the program and plans to continue the program into the future.

References

Abadie, A., and G. Imbens. 2011. "Bias-corrected Matching Estimators for Average Treatment Effects." *Journal of Business and Economic Statistics* 29 (1): 1-11.

Allcott, H. 2011. "Social Norms and Energy Conservation." *Journal of Public Economics* 95: 1082-1095.

American Council for an Energy-Efficient Economy (ACEEE). 2013. "ACEEE Field Guide to Utility-Run Behavior Programs." Prepared by S. Mazur-Stommen and K. Farley. < http://www.aceee.org/research-report/b132>.

Harding, M. and A. Hsiaw. Forthcoming. "Goal Setting and Energy Conservation." *Journal of Economic Behavior and Organization*.

Harding, M. and P. McNamara. 2011. "Rewarding Energy Engagement: Evaluation of Electricity Impacts from CUB Energy Saver, a Residential Efficiency Pilot Program in the ComEd Service Territory." Efficiency 2.0 and Stanford University.

Ho, D.E., K. Imai, G. King, and E. Stuart. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference". *Political Analysis* 15 (3): 199-236.

Imbens, G. and J. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature* 47(1): 5-86.

Navigant Consulting Inc. 2014. "C3-CUB Energy Saver Program EPY5 Evaluation Report." Prepared by B. Provencher and C. McClure. <http://www.ilsag.info/comed_eval_reports.html>.

State and Local Energy Efficiency Action Network (SLEEAN). 2012. "Evaluation, Measurement, and Verification (EM&V) of Residential Behavior-Based Energy Efficiency Programs: Issues and Recommendations." Prepared by A. Todd, E. Stuart, S. Schiller, and C. Goldman, Lawrence Berkeley National Laboratory. <http://behavioranalytics.lbl.gov>.

Appendix

Section 1 – Detailed Model Assumptions[18]

Model 1, model 2, and the regression associated with model 3 are all variations of a linear fixed effect model estimated using panel data. The panel dataset we use has observations of energy usage by households (the individual dimension of the panel) in each month (the time dimension of the panel). Four assumptions ensure that the parameter estimates coming out of a linear fixed effect model are unbiased.

Assumption 1 – Linearity
The model is linear in the parameters, the individual fixed effect, and the error term.

Assumption 2 - Independence
The observations are independent across individuals, but not necessarily across time.

Assumption 3 – Strict Exogeneity
The error term is uncorrelated with any past, present, or future values of the explanatory variables for a given individual. This assumption further implies that the error term is uncorrelated with the individual fixed effect.

Assumption 4 – Identifiability
The explanatory variables are not perfectly collinear, they have non-zero within-variance (i.e. there is variation over time for each individual), and there are relatively few extreme values. This assumption rules out the inclusion of a model intercept or any other time invariant explanatory variable.

For our models, the first assumption is easily verified; simply looking at the models one can see that they are all linear in the necessary terms. Assumption 2 also holds in our sample; we expect that observations of energy usage across households are independent, while observations of energy usage for a single household through time are not. Assumption 4 also holds in our sample; the explanatory variables are not perfectly collinear in any of the three models and there is variation within each individual through time. Assumption 3 must be assumed to hold, as the strict exogeneity assumption is assumed to hold in any ordinary least squares (OLS) model.

We avoid imposing any specific form on the error distribution of our three models and allow for heteroskedasticity and serial correlation of an unknown form within a household by estimating a cluster-robust variance-covariance matrix (Huber, 1967; White, 1980). In this case, the cluster is the household because we expect that the observations are independent across households but correlated within a given household through time (as stated in Assumption 2). Clustering gives a fifth assumption necessary to estimate the cluster-robust variance-covariance matrix.

Assumption 5 – Clustered Errors
$Var(e_i \mid X_i, c_i) = \Omega_{e,i}(X_i)$ is positive definite and independent. Where i indexes the individual cluster, e_i is the error term, X_i is the vector of explanatory variables, c_i is the individual fixed effect, and $\Omega_{e,i}$ is the covariance matrix $\Omega_{e,i} = E(x_i x_i' e_i^2)$. This means that the error terms across different individuals can have different variances and can be correlated within an individual conditional on all the explanatory variables of all the observations within a cluster.

[18] This discussion draws heavily from Schmidheiny (2013a and 2013b) but the assumptions we discuss are standard and can be found in most econometrics textbooks. Schmidheiny provides a list of textbook and other references for those interested in further exploration of these issues.

Under assumptions 1-5 the cluster-robust variance-covariance matrix is estimated by

$$Var(\hat{\beta}\,|\,X) = \sigma^2 (X'X)^{-1}[X'\Omega X](X'X)^{-1}.$$

As the number of clusters, H, goes to infinity, the model parameters, β, are asymptotically normally distributed.

$$\sqrt{H}(\hat{\beta}-\beta) \xrightarrow{d} N(0,\Sigma)$$
$$\Sigma = Q_{XX}^{-1} E(X'ee'X) Q_{XX}^{-1}$$
$$Q_{XX} = E(X'X)$$

Since the cluster-robust variance-covariance matrix imposes no specific form on the error distribution, there is no need to plot the residuals ex-post to verify the assumptions of the model. Under assumptions 1-5, standard t-tests can be performed using large samples, such as the one we have, to verify the statistical significance of the parameters in the models.

Section 2 – Matching on Past Energy Use

The matching literature makes clear that the purpose for matching is to reduce model specification bias. By balancing the treatment and comparison groups on those covariates that have the greatest effect on energy use, estimated program impacts are less sensitive to the model specification. In a setting where the analyst is tempted by conformity bias – the incentive to derive a “favorable” or “expected” result – the advantage of matching the distributions of covariates to generate estimates less sensitive to the model specification is especially salient.

The decision about which covariates to include in matching is circumscribed by the available data, and the principle that matching should be limited to only the most important covariates. Matching on too many variables reduces the quality of the match for its designated purpose –controlling for model specification bias – by causing worse matching on the covariates that have the greatest explanatory power in the subsequent regression model.

The good news for analysts applying matching methods to energy analysis is that the most important variable for predicting energy use is also typically the variable most readily available: past energy use. In past studies these authors have found that, including past energy use as a predictor causes almost all demographic and structural covariates to become statistically non-significant. The implication is that, from the perspective of mitigating model specification bias, it is not important to match on these other variables, and, following the logic presented above, matching on these other variables can actually reduce the advantage of matching by reducing the quality of the match of the dominant variable.

Additionally, if the analyst believes that other variables are important to the regression, covariates which are not matched upon can still be added into a post-matching regression model if the data is available. This mitigates the problem of limiting the potential matches by matching on too many covariates but still allows the analyst to control for potentially important demographic factors.

Finally, although matching on past energy use is not designed to address selection bias, it has the effect of mitigating selection bias to the extent that the unobserved variables embedded in the regression disturbance term that are correlated with program enrollment are also correlated with past energy use – a prospect that would seem likely (though there is no claim that this corrective effect is any greater than simply including past energy use in a standard regression analysis without matching).

Section 3 – References

Huber, P.J. 1967. “The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions.” *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1: 221-233.

Schmidheiny, Kurt. 2013. “Clustering in the Linear Model.” *Short Guides to Econometrics.* Unversität Basel. <http://kurt.schmidheiny.name/teaching/clustering.pdf>.

Schmidheiny, Kurt. 2013. “Panel Data: Fixed and Random Effects.” *Short Guides to Econometrics.* Unversität Basel. <http://www.schmidheiny.name/teaching/panel.pdf>.

White, H. 1980. “A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity.” *Econometrica* 48: 817-838.