Matching and VIA: Quasi-Experimental Methods in a World of Imperfect Data

Eileen Hannigan, Illume Advising, LLC, Madison, WI
Jan Cook, Accelerated Innovations, St. Paul, MN

ABSTRACT

The State and Local Energy Efficiency (SEE) Action Network strongly encourages the use of random control trials (RCT) in evaluations of behavior change programs. For situations where RCT is not possible, SEE Action reviews and ranks several quasi-experimental methods. In descending rank order, the report discusses regression discontinuity, variation in adoption (VIA), and matched control groups. Program and design considerations often restrict impact analysis methods to quasi-experimental designs. Options are further restricted by the particular requirements of the quasi-experimental designs. Consequently, for many programs RCT and regression discontinuity, the two evaluation methods most highly rated by SEE Action are not possible to implement, leaving VIA and matching as possible evaluation methods.

While matching and VIA also have restrictive assumptions, certain multi-year programs are compatible with both types of analysis, providing an opportunity to assess the approaches by comparing the execution of and results from each method. In this paper we use a longitudinal data set of energy use and program participation from an opt-in behavior program that was deployed by four utilities over six years. We compare impact estimates of the program that were derived using a matching method to those derived from using a VIA approach. We also explore the strengths and limitations of each method, the extent to which the data fulfill the assumptions of each method, and the practical consequences of choosing one method over the other. In addition, we discuss the sensitivity of each method to sample size, variability, and program-specific deployment and treatment conditions.

Introduction

To estimate the impact of an intervention of any kind, the post-intervention outcome of interest must be compared to a counterfactual – what would have happened in the absence of the program. Since we are unable to know exactly what would have happened to the same group of treatment customers in the absence of the intervention, we need to develop a counterfactual that is a proxy for the treatment group without intervention scenario. In randomized control trials, the counterfactual is supplied by the randomly-assigned control group. Successful random assignment assures that the only difference between the treatment and control groups is whether or not they received the treatment.

However, there are many cases where randomized control trials are not possible. For example, sample sizes in small jurisdictions limit the ability to retain control groups and meet savings goals. Second, the program may utilize mass media and marketing approaches that are not conducive to experimental models. In addition, a randomized control trial requires denying access to customers who may want to participate in a program which may not be acceptable to stakeholders. Finally, if an evaluation is conducted after the program is implemented in-field, an experimental design cannot be established or, even if an experimental design was implemented, there may be errors in the randomization that negates the design.

For situations where a randomized control design is not possible, SEE Action reviews and ranks several quasi-experimental methods (SEE Action 2012). In descending rank order the report discusses regression discontinuity design, variation in adoption (VIA) and matched control groups. As noted, program and design considerations often restrict impact analysis methods to quasi-experimental designs. Options are further restricted by the particular requirements of the quasi-experimental designs. For example, regression discontinuity can be applied only if the eligibility requirement for households to participate in a program is a cutoff value of a characteristic that varies within the population. Consequently, for many programs, the two methods for establishing a counterfactual most highly rated by SEE Action are not possible to implement, leaving VIA and matched control groups as possible evaluation methods.

In this paper we use a longitudinal data set of energy use and program participation from an opt-in behavior program that was deployed by four utilities over six years. We compare impact estimates of the program that were derived using a matching method to those derived from using a VIA approach. We also explore the strengths and limitations of each method, detail the extent to which the data fulfill the assumptions of each method, and discuss the practical consequences of choosing one method over the other. In addition, we discuss the sensitivity of each method to sample size, variability, and program-specific deployment and treatment conditions.

Intervention Overview

To compare approaches and results using both VIA and matched control group methods we use a longitudinal data set of energy use and program participation from an opt-in behavior program that was deployed by four utilities over six years. The program, called MyMeter, is an online platform that provides various services to utilities and their customers. The platform helps utilities operate programs that manage customer usage loads and achieve energy savings. Capabilities include dynamic pricing programs; air conditioning cycling for residential and small business customers; direct load control programs for large commercial and industrial customers; and behavioral energy efficiency programs. MyMeter can also run contests and challenges that promote energy conservation.

MyMeter provides visualization tools that enable customers to track energy usage and billing information. These tools include:

- Comparative usage: a feature that benchmarks customer usage against their own usage history and others in the territory.
- Energy challenges: customers set their own conservation goals and track their progress.
- Property profile: customers fill out detailed information on their homes and businesses.
- Bill threshold alerts: notifies customers when they hit pre-set usage thresholds.
- Peak time alerts: notifies customers when peak demand hours are occurring.
- Energy markers: tracks major changes in the home that may impact usage.
- Outage alerts: notifies customers about power outages in their region.

These features are available on desktop and mobile devices and allow customers to see how their usage changes over time; how weather, occupancy, and appliance use affect their usage patterns; and how they compare to their neighbors.

Customers enroll in the program by logging on to a web portal and creating an account. Any level of activity qualifies a customer as an enrolled participant. Almost half of the participants logged in between 2 and 9 times and 17 percent logged in 10 or more times. About 6 percent also used at least one additional feature such as energy markers, property profiles, or threshold alerts.

Methods

We applied both the VIA and matched control group method to data from the MyMeter program to compare the applicability and outcomes of each method.

Variation in Adoption

VIA takes advantage of the temporally staggered sign up rates of opt-in programs by comparing households that have opted in to the program to households that have not yet opted in, but will do so at a later date. The later adopters are the control group for the earlier adopters. Data for non-participants is not needed. However, the model is based on the assumption that the only difference between earlier and later adopters is awareness of the program. Ideally, the decision by each household to opt-in at a particular time is effectively random, influenced only by exposure to marketing or other communication that makes participants aware of the program.

To date, VIA has not been used widely in published energy efficiency program evaluation. The method was introduced by Harding and Hsiaw (2012) in a study on the effect of goal-setting on energy savings. The method was also applied to a web-based opt-in behavioral program in Illinois (Provencher & McClure 2014) and was reviewed in a 2013 IEPEC paper (Glinnsman & Provencher 2013).

The VIA model can be specified with a customer-specific fixed effect and three sets of binary variables:

- Monthly variation in energy usage independent of the program (e.g. effects of weather) is accounted for with a set of binary variables indicating the calendar month.
- Energy consumption before program enrollment is controlled for with a set of binary variables for the number of months until program enrollment.
- A set of binary variables indicating the number of months since program enrollment allows program impacts to vary by time in the program.

Formally, the regression model is:

$$kWh_{it} = \alpha_i + \sum_{m=1}^{M} \beta_m \times Month_{mt} + \sum_{r=1}^{R} \gamma_r \times PrePeriodMonth_{rit} + \sum_{p=1}^{P} \delta_p \times PostPeriodMonth_{pit} + \epsilon_{it}$$

Where,

kWh_{it}	= average daily energy use by customer *i* in billing period *t*.
α_i	= household-specific constant.
$Month_{mt}$	= binary variables indicating the calendar month of the billing period.
$PrePeriodMonth_{rit}$	= binary variables indicating the number of months until program enrollment for customer *i* during billing period *t*.
$PostPeriodMonth_{pit}$	= binary variables indicating the number of months since program enrollment for customer *i* during billing period *t*.
ϵ_{it}	= model error term.
$\beta_m,\ \delta_p,\ \gamma_r$	= model coefficients to be estimated.

Per participant annual savings across an entire year of participation results from summing δ_p, the post-period month parameters.

Matched Control Group Method

While VIA uses later adopters as a counterfactual to earlier adopters, the matched control group method establishes a counterfactual using non-participants. The method constructs a non-random control group made up of households that are as similar to the treatment group as possible on covariates that have a high correlation with the outcome variable. In the case of evaluating a program's impact on energy use, the outcome variable is monthly post-program energy use and the variable with the greatest correlation with this outcome is energy use in the same month of the pre-program period. To construct a control group, pre-program energy usage for both participants and a large pool of non-participants is needed. A matched control household is selected for each participating household. Selection is based on the minimum sum of squared differences in monthly energy usage in the pre-period. Matches can be selected with or without replacement. With replacement means that a control household can be selected as a match for more than one participating household. The matched control group method has been described in some detail in the academic literature (Abadie & Imbens 2011; Provencher et al 2013; Stuart 2010). This method has been used and approved by utility stakeholder or regulatory groups in several jurisdictions including Massachusetts, Rhode Island, and Minnesota (Opinion Dynamics 2013; Illume Advising and Klos Energy 2014; Illume Advising and Navigant Consulting 2014).

Of the four utilities included in this analysis, two had significant numbers of seasonal residents with large variations in usage. We incorporated usage variation in the matching so that participants were matched to non-participants with similar levels of variability. We used one to one matching with replacement.

We estimated separate models for each utility. The regression models are applied only to the post-enrollment periods. Formally the regression model is:

$$kWH_{kt} = \alpha_{0t} + \beta \times Partic_{kt} + \gamma \times PrekWh_{kt} + \varepsilon_{kt}$$

Where,

kWH_{kt}	= the average daily electricity use by household k in month t of the post-period.
α_{0t}	= a monthly fixed effect.
$Partic_{kt}$	= a binary variable with a value of 1 for participants and 0 for matched non-participants.
$PrekWh_{kt}$	= the average daily pre-participation electricity use by household k that is also the same calendar month as month t.
ε_{kt}	= model error term.

Results

We compared the application of both VIA and matched control group methods to the longitudinal data set of program participants with particular attention to enrollment timing, enrollment saturation, selection bias, savings estimates, and program-specific conditions.

Enrollment Timing

As noted above, VIA assumes that households opt-in at different points in time. Consequently, a program must have enough variation in enrollment dates to have adequate periods of earlier and later participation. Given the length of time the MyMeter program has been implemented, we were able to select customers with at least 12 months of program participation data. Figure 1 shows the distribution of enrollments by year and month.

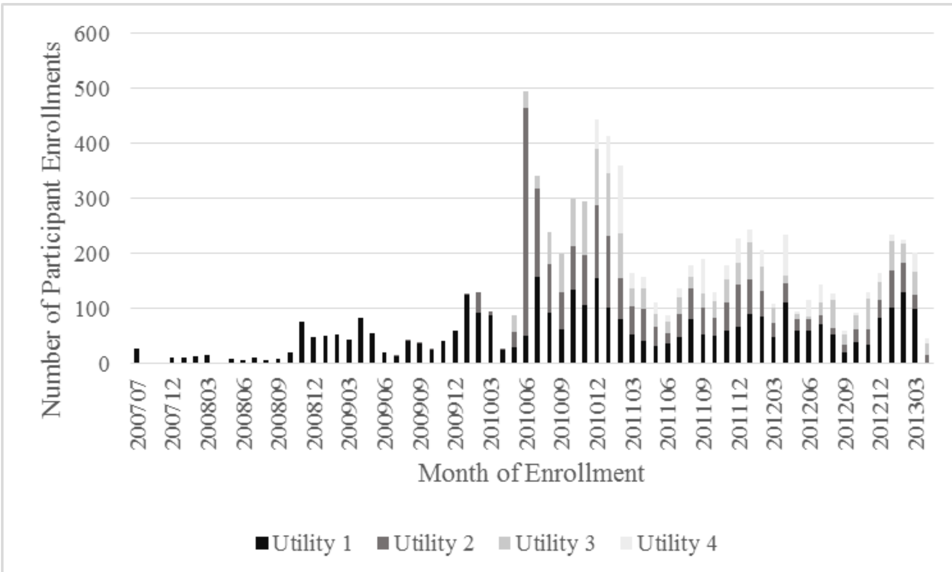

Figure 1. Distribution of Program Enrollments by Utility

Since energy use varies seasonally, the best implementation of the matched control group method draws on at least 12 months of pre-enrollment usage data for establishing the matched comparison group and at least 12 months of post-enrollment data for estimating savings. (With either method it is not valid to extrapolate savings from one part of the year to another part of the year.) For this analysis, we had ample months of pre- and post-data to establish matches and estimate savings.

Enrollment Saturation

One strength of VIA is that there is no limit on the portion of the customer base that enrolls. Since later enrollees are the counterfactual for earlier enrollees, it is theoretically fine if all potential participants eventually enroll as long as their enrollment is distributed over time.

The matched control group does not strictly prohibit any portion of the customer base from enrolling as would happen with a randomized control trial, however, a sufficient number of customers are needed in the non-participant group to find adequate matches for the participants. For this analysis our pool of non-participating customers was 5 to 10 times greater than our pool of participants. In addition, the quality of the matches was high, as shown in Figure 3.

Selection Bias

Both VIA and the matched control group method can be sensitive to selection bias. VIA assumes that the only difference between earlier and later adopters is knowledge of the program. This assumption is tested in the model specification with the binary variables indicating months until program adoption. If the assumptions are upheld, then savings prior to program enrollment should be statistically equivalent to zero since the customer has not yet enrolled in the program. If customers are more likely to reduce energy use as they approach enrollment, then the pre-program parameters will be different from zero. If customers are not reducing their energy use prior to enrollment then γ_r will equal zero for all r. We found that for each utility between four and eight of the 12 pre-enrollment parameters were statistically different from zero meaning that the model assumptions could be violated and results should be interpreted with caution. Figure 2 shows pre-program enrollment average daily kWh usage by year of program enrollment. For the utilities shown, average usage among later enrollees is lower than average usage among earlier enrollees.

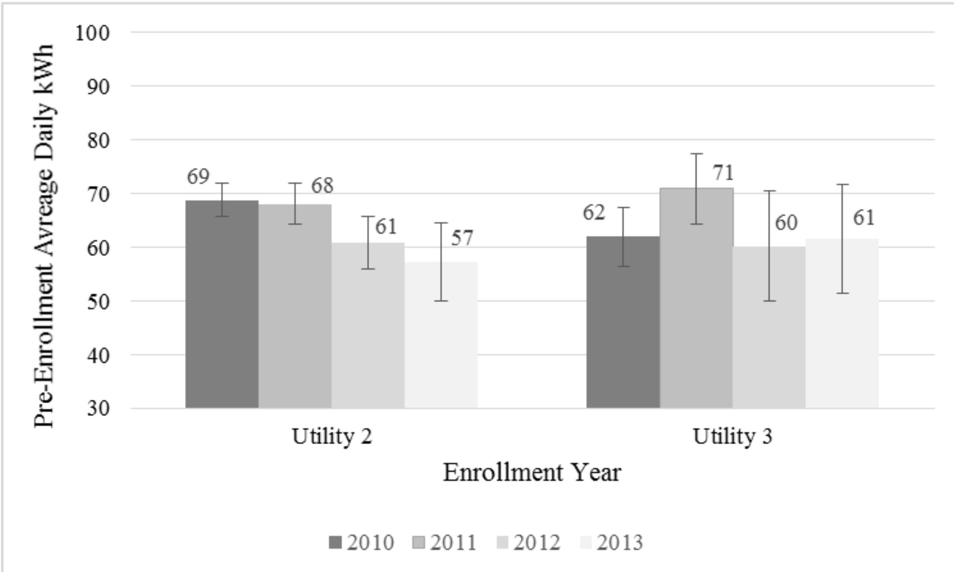

Figure 2. Comparison of Average Daily kWh by Year of Enrollment

The matched control group method assumes all differences between participants and comparison customers are captured in their pre-enrollment energy usage.[1] Consequently, selection bias is a concern with this method. While there is no statistical test for selection bias, Imbens and Wooldridge (2009) present a quasi-test. Applied to an opt-in energy program, the logic of the test is that if there is no 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.

One way to implement the test is to match customers based on a year of pre-period usage starting 16 months prior to the program through 5 months prior to the program. Then months 1 through 4 prior to the program are used as a test period. If the difference in energy use between the treatment and comparison groups in the test period is not statistically or practically different from zero then the test boosts confidence that selection bias is not a critical issue in the analysis. If the treatment group was more likely to do something to save energy even in the absence of the intervention, then we would expect to see divergence in energy use between the treatment and comparison groups in the four months between the matching period and the program period.

Figure 3 displays the average daily energy use for treatment and comparison customers across a 20-month pre-enrollment period. Average usage remains very similar through the months up to enrollment.

[1] Matching algorithms can be enhanced with other data such demographics, but these are rarely available.

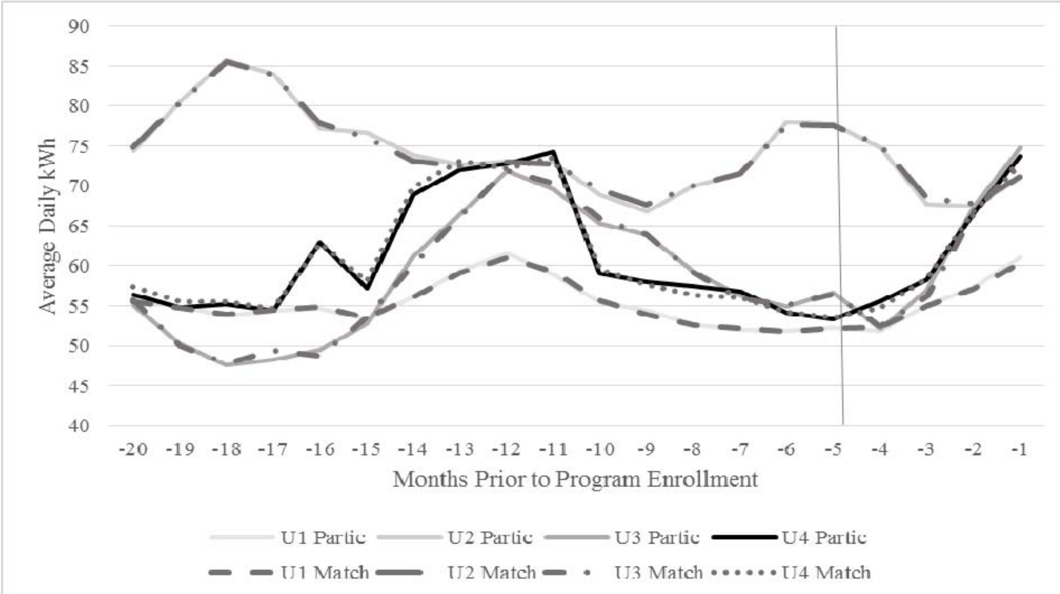

Figure 3. Comparison of Average Daily kWh for Months Prior to Enrollment

Savings Estimates

Figure 4 compares the savings estimate for each utility for each method. For three out of four utilities the matched control group method resulted in lower savings estimates with smaller confidence intervals. The VIA method produced savings estimates that were statistically different from zero for only one utility. As noted, the data do not fulfill one of the assumption of VIA as the earlier enrollees differed from later enrollees on pre-enrollment usage. Notably, in utilities one through three, the VIA savings estimate falls within the 90% confidence interval of the matched control group estimate.

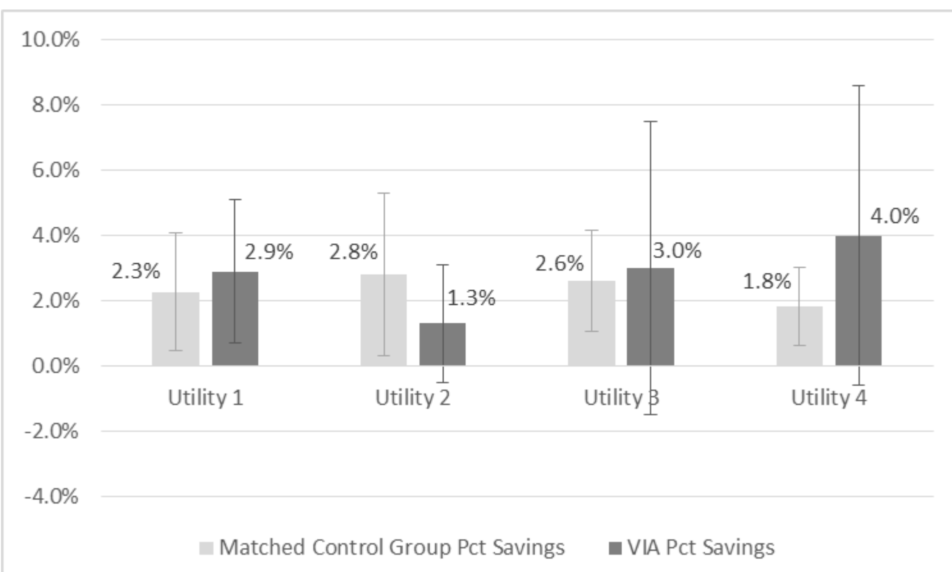

Figure 4. Average Annual Percentage Savings and 90% Confidence Intervals by Utility and Method

Conclusions

The evaluation analyst’s world often consists of imperfect data from non-randomized studies. Various considerations often prevent the implementation of a randomized control trial necessitating the use of quasi-experimental methods. We reviewed two methods that can work well for opt-in behavior programs with rolling enrollment: VIA and matched comparison groups. The key questions and considerations for choosing method are summarized in Table 1. VIA allows the program to enroll all interested participants as long as enrollment is distributed across time; has a built-in test of assumptions; and only requires billing data from participants. However, VIA requires that earlier enrollees are similar to later enrollees. This proved to be a problem for this particular program as later enrollees had lower pre-enrollment energy use than earlier enrollees. Differences between earlier and later enrollments is quite common. Often utilities specifically target one type of customer first and then roll out the program to other customers based on geography, usage, and other factors. For a VIA method to be effective, the program roll-out needs to be planned carefully with the VIA evaluation approach in mind.

For the matched control group method, the analyst selects non-participants that have similar pre-enrollment usage to participants, and estimates impacts using a simpler regression model. The matched comparison method resulted in more conservative savings estimates with smaller confidence intervals in three out of four utilities. The key requirement for a successful matched comparison group analysis is the availability of a pool of non-participants with adequate billing history from which to draw matches. Unlike randomized designs and VIA, the matched control group method can be implemented even in situations where evaluation methodology has not been considered until after the program is in the field.

For either method, it is important to take into consideration the program-specific conditions. For example, two of the utilities included in this analysis have a significant portion of seasonal residents in their territories. This complicated finding matched control customers. In addition, enrollments in the program spanned several years. As a result, customers who enrolled later may have been exposed to program marketing and other influences on their energy use for some period of time before enrolling, violating the condition that later enrollees only differ from earlier enrollees by knowledge of the program. In many practical applications this condition may be difficult to uphold, particularly for programs implemented in smaller communities and communicated through broad channels. VIA is a good method, but has some restrictive assumptions that require a specific set of program conditions that are not often possible or practical. While the matched control group method has limitations, it better accommodates the typical and practical considerations of program planning and implementation.

Table 1. Key Questions and Considerations for VIA and Matched Control Group Method

Factor	**VIA**	**Matched Control Group Method (MCGM)**
Selection Bias	Are customers who enroll later similar to customers who enroll earlier except for knowledge of the program? VIA model provides built-in test of this assumption. All customers in the model have opted in, limiting selection bias.	Does matching on energy usage control for other differences between participants and comparison customers? There is no test, but matching on 16 to 5 months prior to enrollment and using month 4 to month 1 prior to enrollment as a test period is suggestive.
Enrollment Saturation	Are there enough earlier and later adopters?	Is there a large enough pool of non-participants to draw from for matches?
Enrollment Timing	Was enrollment spread across months or a year?	Are 12 months of pre- and post-enrollment usage data available?
Territory- and program-specific conditions	Can communication about and knowledge of the program be restricted to particular groups of	Are there specific customer types in the territory that have unusual energy usage patterns that may be difficult to match?

	customers at different points in time?	
Evaluation Planning	VIA requires planning for evaluation before program implementation to ensure model assumptions are met	Model assumptions are less restrictive and approach can be applied after program implementation including when there was a randomization failure in an RCT.
Data Availability	Are energy usage data available for earlier and later enrollees?	Are pre- and post-enrollment usage data available for enrollees and a large pool of non-participating customers?
Industry Experience	Currently there are limited examples of VIA use in the industry.	MCGM has been used in several jurisdictions.

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.

Glinnsman, B. and B. Provencher. 2013. I can't use a Randomized Controlled Trial – NOW WHAT? Comparison of Methods for Assessing Impacts from Opt-In Behavioral Programs. Paper presented at the 2013 International Energy Program Evaluation Conference. August 13-15.

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

Illume Advising and Klos Energy. March 2014. MyMeter Multi-Utility Impact Findings. Prepared for Accelerated Innovations.

Illume Advising and Navigant Consulting. 2014. Rhode Island Behavioral Program and Pilot Process and Impact Evaluation. Prepared for National Grid Rhode Island.

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

Opinion Dynamics. June 2013. Massachusetts Cross-Cutting Behavioral Program Evaluation Integrated Report. Delivered to Massachusetts Energy Efficiency Advisory Council & Behavioral Research Team.

Provencher, B., B. Vittetoe-Glinsmann, A. Dougherty, K. Randazzo, P. Moffittd, and R. Prahl. 2013. Some Insights on Matching Methods in Estimating Energy Savings for an Opt-In, Behavioral-Based Energy Efficiency Program. Paper presented at the 2013 International Energy Program Evaluation Conference. August 13-15.

Provencher, B. and C. McClure. January 2014. C3-CUB Energy Saver Program EPY5 Evaluation Report Final. Presented to Commonwealth Edison Company, Oakbrook Terrace, IL.

State and Local Energy Efficiency Action Network. 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.

Stuart, E.A. 2010. “Matching methods for causal inference: A review and a look forward.” Statistical Science, 25 (1): 1-21.