NAVIGANT

Home Energy Reports Program Evaluation Report

FINAL

Energy Efficiency Plan: Gas Plan Year 4 (6/1/2014-5/31/2015)

Presented to Peoples Gas and North Shore Gas

March 1, 2016

Prepared by:

Paul Higgins, Daniel Layton, and Anusha Jagannathan
Navigant Consulting, Inc.

www.navigant.com

NAVIGANT

Submitted to:

Peoples Gas
North Shore Gas
200 East Randolph Street
Chicago, IL 60601

Submitted by:

Navigant Consulting, Inc.
30 S. Wacker Drive, Suite 3100
Chicago, IL 60606

Contact:

Randy Gunn, Managing Director
312.583.5714
Randy.Gunn@Navigant.com

Kevin Grabner, Assoc. Dir.
608.497.2327
Kevin.Grabner@Navigant.com

Robert Neumann, Assoc. Dir.
312.583.2176
Rob.Neumann@Navigant.com

NAVIGANT

Table of Contents

E. Executive Summary 3

E.1. Program Savings 4
E.2. Program Volumetric Detail 5
E.3. Findings and Recommendations 5

1 Introduction 7

1.1 Program Description 7
1.2 Evaluation Objectives 8

2 Evaluation Approach 9

2.1 Overview of Data Collection Activities 9
2.2 Sampling Plan 9
2.3 Data Used in Impact Analysis 9
2.4 Statistical Models Used in the Impact Evaluation 10
2.5 Accounting for Uplift in Other Energy Efficiency Programs 11
2.6 Process Evaluation 12

3 Gross Impact Evaluation 13

3.1 PPR and LFER Model Parameter Estimates 14
3.2 Uplift of Savings in Other Energy Efficiency Programs 16

4 Net Impact Evaluation 18

5 Findings and Recommendations 19

6 Appendix 21

6.1 Detailed Impact Methodology 21
6.1.1 PPR Model 21
6.1.2 LFER Model 22
6.1.3 Detailed Impact Results: Parameter Estimates 22

NAVIGANT

List of Figures and Tables

Figures

Figure 3-1. Comparison of PPR and LFER Estimates of PG GPY4 HER Savings 14
Figure 3-2. Comparison of PPR and LFER Estimates of NSG GPY4 HER Savings 15

Tables

Table E-1. GPY4 Peoples Gas HER Program Natural Gas Savings 4
Table E-2. GPY4 North Shore Gas HER Program Natural Gas Savings 4
Table E-3. GPY4 Peoples Gas Home Energy Reports Program Participation Detail 5
Table E-4. GPY4 North Shore Gas Home Energy Reports Program Participation Detail 5
Table 1-1. Synopsis of PG and NSG Programs 8
Table 2-1. Primary Data Collection Activities 9
Table 2-2. Impact of Data Cleaning on Analysis Dataset 10
Table 3-1. GPY4 PG Total HER Program Savings 13
Table 3-2. GPY4 NSG Total HER Program Savings 14
Table 3-3. GPY4 PG HER Uplift Adjustment Details 16
Table 3-4. GPY4 NSG HER Uplift Adjustment Details 17
Table 6-1. Savings Parameter Estimates 22

E. Executive Summary

This report summarizes Navigant Consulting, Inc.'s (Navigant's) findings and results from the impact and process evaluation of the fourth program year (GPY4)[1] of the Peoples Gas (PG) and North Shore Gas (NSG) Home Energy Reports (HER) programs. Initially launched in 2013[2], these programs are designed to generate energy savings by providing residential customers with information about their energy use and energy conservation suggestions and tips. Program participants receive information in the form of home energy reports that give customers various types of information, including the following:

- Assessment of how their recent energy use compares to their own energy use in the past
- Tips on how to reduce energy consumption, some of which are tailored to the their own circumstances
- Information on how their energy use compares to that of neighbors with similar homes

Recipient customers receive several reports per year by mail and also have the option to log onto a dedicated program website to learn more ways to save energy and to report conservation steps they have taken. Other studies have shown that receiving reports containing this information can stimulate customers to reduce their energy use, creating average energy savings in the one percent to three percent range, depending on local energy use patterns.

An important feature of the PG and NSG HER programs is that both are designed as randomized controlled trials (RCTs).[3] Customers in the target group of residential customers from each utility are randomly assigned to either the recipient group or the control (non-recipient) group for the purpose of estimating changes in energy use due to the program. Customers may opt *out* of the program at any time, however, they cannot opt *in* due to the RCT design. An implication of the RCT design is that the savings estimates are intrinsically net of free-ridership and most spillover bias.

In its evaluation of the GPY3 PG HER program, Navigant verified net savings of 2,072,182 therms before uplift adjustment[4], representing an average recipient savings rate of 0.85 percent. The verified net savings after uplift adjustment for the PG program in GPY3 was 2,054,727 therms.

In its evaluation of the GPY3 NSG HER program, Navigant verified net savings of 662,518 therms before uplift adjustment, representing an average recipient savings rate of 0.62 percent. The verified net savings after uplift adjustment for the NSG program in GPY3 was 652,718 therms.

The design of the program did not change in GPY4.[5]

---

[1] GPY4 began June 1, 2014 and ended May 31, 2015.

[2] The HER programs were initially rolled out to targeted samples of 151,200 PG customers and 91,350 NSG customers beginning in October 2013. Control groups of approximately 21,000 were also selected for each program.

[3] The program implementer, Opower, randomly allocated targeted PG and NSG residential customers between participant and control groups. As part of its GPY3 impact evaluation, Navigant confirmed that the usage data were consistent with an RCT design.

[4] Uplift refers to the impact of the HER program on enrollment in *other* EE programs. To avoid double-counting savings, Navigant subtracts estimated uplift savings from the total HER program savings.

[5] Note that the GPY3 programs were not rolled out until October 2013; thus, HER recipients were not exposed to the program for a full twelve months in GPY3. See GPY3 PG-NSG Home Energy Reports Program Evaluation Report, which is available at http://www.ilsag.info/evaluation-documents.html, for more information.

E.1. Program Savings

Table E-1 summarizes the GPY4 natural gas savings from the Peoples Gas HER Program.

Table E-1. GPY4 Peoples Gas HER Program Natural Gas Savings

Program/Path	Ex Ante Savings[6] (Therms)*	Verified Savings Prior to Uplift Adjustment (Therms)	Verified Realization Rate[7]	Total Uplift Adjustment (Therms)	Verified Savings After Uplift Adjustment (Therms)
Home Energy Reports	3,009,588	3,318,421	1.10	37,981	3,280,440

Source: Navigant analysis of GPY4 customer billing data and program tracking data (July 20, 2015 data extract).
* Navigant telephone communication with Franklin Energy in September 2015.

Table E-2 summarizes the natural gas savings from the GPY4 North Shore Gas HER Program.

Table E-2. GPY4 North Shore Gas HER Program Natural Gas Savings

Program/Path	Ex Ante Savings (Therms)*	Verified Savings Prior to Uplift Adjustment (Therms)	Verified Realization Rate	Total Uplift Adjustment (Therms)	Verified Savings After Uplift Adjustment (Therms)
Home Energy Reports	874,691	1,094,406	1.25	-14,159	1,108,565

Source: Navigant analysis of GPY4 customer billing data and program tracking data (July 20, 2015 data extract).
* Navigant telephone communication with Franklin Energy in September 2015.

The PG HER Program had a realization rate of 1.10 in GPY4, indicating that the verified savings was ten percent higher than the level expected. The NSG HER Program achieved a realization rate of 1.25, indicating that verified savings was 25 percent higher than had been anticipated. We review this result in greater detail in Section 3.

[6] The term "Ex Ante" refers to the forecasted savings reported by the Program Administrator that have not been independently verified through evaluation. Savings that have been independently verified by Navigant are referred to as "Verified".

[7] Verified Gross Realization Rate (RR) = Verified Gross Savings/Ex Ante Gross Savings.
Verified Gross Savings = RR * Ex Ante Gross Savings

E.2. Program Volumetric Detail

Table E-3 and Table E-4 below present GPY4 program participation reported by the Program Administrator Franklin Energy Services (FES) for the Peoples Gas and North Shore Gas programs.

Table E-3. GPY4 Peoples Gas Home Energy Reports Program Participation Detail

Participation	Program Total
Targeted Number of Participants	151,200
Targeted Number of Controls	21,000
Average daily savings per participant in GPY4 (therms)	0.06
(Standard error)	*(0.007)*
Average savings rate in GPY4 (percent)	1.19

Source: Navigant analysis of GPY4 customer billing data and program tracking data (July 20, 2015 data extract).

Table E-4. GPY4 North Shore Gas Home Energy Reports Program Participation Detail

Participation	Program Total
Targeted Number of Participants	91,350
Targeted Number of Controls	21,000
Average daily savings per participant in GPY4 (therms)	0.03
(Standard error)	*(0.005)*
Average savings rate in GPY4 (percent)	0.88

Source: Navigant analysis of GPY4 customer billing data and program tracking data (July 20, 2015 data extract).

E.3. Findings and Recommendations

The following provides insight into key program findings and recommendations.

Verified Net Savings, Gross Savings and Realization Rate.

Finding 1. Overall, both programs improved their savings performance in GPY4 relative to GPY3. In GPY4 the PG HER Program achieved verified net savings of 3,318,421 therms prior to uplift adjustment, which represents a savings rate of 1.19 percent. In GPY3, the corresponding savings was 2,072,182 therms, or 0.85 percent. In PGY4 the NSG program achieved verified savings of 1,094,406 therms prior to uplift adjustment, which represents a savings rate of 0.88 percent. In GPY3, the corresponding figure was 662,518 therms, or 0.62 percent.

Finding 2. The therms savings reported in Finding 1 are not strictly comparable across program years because the PG and NSG HER program recipients did not start receiving reports until October 2013 and thus were exposed to the program's influence for only eight months of GPY3. Additionally, the numbers of participants in GPY4 were somewhat reduced from those prevailing at the rollout of the program in GPY3 due to customer drop-outs and move-outs. The more apt comparison between program years is on the basis of the percentage savings rates, which improved year-over-year for both programs.

Finding 3. Opower indicated in a discussion with the evaluation team in 2015[8] that they intended to roll out a second wave of approximately 12,000 new participants in the NSG service territory, along with approximately 10,000 new controls. They said they expected to start sending out reports to the new wave of participants before the end of 2015.

Recommendation 1. In view of the strong performance of the PG and NSG HER programs in GPY4 – in particular the NSG program's 1.25 realization rate – and the advent of a new wave of recipients in GPY5, Navigant recommends that both utilities consider undertaking persistence research on their HER programs beginning in GPY6. This research, which would entail randomly selecting a subset of recipient customers whose reports would be terminated, would permit evaluators to measure the rate at which HER program savings rates decay once reports cease being delivered, and thus provide insight into persistence and measure life.

[8] Telephone interview with Opower project managers, May 19, 2015.

1 Introduction

1.1 Program Description

This report presents a summary of the findings and results from the impact and process evaluation of the GPY4 Peoples Gas (PG) and North Shore Gas (NSG) Home Energy Reports (HER) programs. These programs are designed to generate energy savings by providing residential customers with information about their energy use and energy conservation suggestions and tips. Program participants receive information in the form of home energy reports that give customers various types of information, including the following:

- Assessment of how their recent energy use compares to their own energy use in the past
- Tips on how to reduce energy consumption, some of which are tailored to the their own circumstances
- Information on how their energy use compares to that of neighbors with similar homes

In GPY4, recipient customers received four reports by mail.[9] They were also invited to log onto a dedicated program website that offers suggestions of additional opportunities to save energy, and also allows participants to fine-tune their customer profiles and report conservation steps that they have taken. Other studies have shown that receiving reports containing this type of information can stimulate customers to reduce their energy use, creating average energy savings in the one percent to three percent range, depending on local energy use patterns.

An important feature of the PG and NSG HER programs is that both are designed as randomized controlled trials (RCTs). Customers in the target group of residential customers from each utility were randomly assigned to either the recipient group or the control (non-recipient) group for the purpose of estimating changes in energy use due to the program. This approach makes the process of verifying energy savings much simpler and more robust: among other things it effectively eliminates free-ridership bias and thus the need for net-to-gross research. Customers may opt *out* of the program at any time, but they cannot opt *in* due to the RCT design.

Working with the program implementation contractor, Opower, the utilities rolled out their HER programs to targeted samples of 151,200 PG customers and 91,350 NSG customers, respectively, with the initial reports going out in October 2013. Control groups of approximately 21,000 customers were also selected for each utility at that time. These are summarized in Table 1-1.

[9] To boost savings in the NSG territory, which has had a persistently lower savings rate than the PG program to date, the program added a fifth report and also implemented a door-hanger campaign in GPY5. (Telephone interview with Opower program managers, *op. cit.*)

Table 1-1. Synopsis of PG and NSG Programs

Utility	Month of First Report	Month of Last GPY4 Report	Targeted Number of Participants*	Targeted Number of Controls*	Average Daily Usage in Post Period (Therms)
PG	October 2013	May 2015	151,200	21,000	5.03
NSG	October 2013	May 2015	91,350	21,000	3.76

Source: Navigant analysis of PG and NSG HER program tracking and customer billing data.

* These are the targeted numbers of customers in each group. Navigant's evaluation analysis used the actual numbers of recipients and control customers in the programs at the start of GPY4.

In its GPY3 evaluation report, Navigant confirmed the RCT design of both utility's programs by comparing the distributions of monthly energy usage of each treatment group-control group pair and verifying that they were consistent with randomized allocation. Since no significant changes were made to the design of either program in GPY4, we did not repeat this exercise in GPY4.[10]

1.2 Evaluation Objectives

The primary objective of the analyses described in this report was to determine the extent to which participants in the PG and NSG HER programs reduced their energy consumption in GPY4 due to the program, and to assess how program savings changed from the previous program year. A secondary objective was identifying uplift in other PG and NSG energy efficiency (EE) programs due to the Opower programs to avoid double-counting of energy savings. The only process research Navigant pursued for either program in GPY4 consisted of interviewing the program managers, so our ability to address questions such as why verified savings differed from ex ante savings or why the savings rates differed between PG and NSG was limited.

[10] See PG-NSG Home Energy Reports GPY3 Evaluation Report, *op. cit.*

2 Evaluation Approach

The evaluation approach used to produce the results presented in this report is consistent with that of the evaluation in the previous program year, and with evaluations of similar programs in other utilities' territories, relying on statistical analysis appropriate for measuring the impacts of RCTs.

2.1 Overview of Data Collection Activities

Navigant received tracking and monthly billing data for all program participants and control customers for the June 2014 to May 2015 period from the program implementer, as shown in Table 2-1.

Table 2-1. Primary Data Collection Activities

Collection Method	Subject Data	Quantity	Net Impact	Process
Customer Billing Data	Program Participants and Controls	All	X	N/A
Program Tracking Data	Program Participants and Controls	All	X	N/A
Tracking Data for Other Programs	Participants in Other Programs	All	X	N/A

Source: Navigant analysis.

For purposes of estimating the GPY4 program impacts, besides the above data Navigant also used the year of pre-program billing data on all program participants and controls obtained as part of its evaluation of the GPY3 impacts.

2.2 Sampling Plan

The PG and NSG HER programs were implemented by the program implementer as an RCT, in which individual customers from each utility's target customer group were randomly assigned to either a treatment (participant) group or a control group for the purpose of measuring the energy savings attributable to the program. Data for all participants and controls are included in the impact evaluation.

2.3 Data Used in Impact Analysis

In preparation for the impact evaluation, Navigant combined and cleaned the data provided by the implementer. The dataset included 284,548 customers: 242,549 program participants and 41,999 controls. Navigant performed the following data cleaning steps:

- Exclude data from outside of the period of examination (June 2014 to May 2015)
- Exclude bills where number of bill days is zero
- Exclude duplicate bills
- Exclude bills with negative usage
- Exclude observations where number of days in bill period is > 40 or < 20
- Exclude outlier observations, defined as observations with average daily usage greater than one order of magnitude from the median usage[11]

[11] Medians were calculated separately by utility. The medians were 3.933 therms per day for PG and 2.6486 therms per day for NSG. Navigant excluded from the analysis all observations in each utility dataset with usage values greater than 10 times the median therms per day or less than the median divided by 10 therms per day.

The impact of each data cleaning step on the numbers of customers and observations remaining in the analysis dataset are shown in Table 2-2.

Table 2-2. Impact of Data Cleaning on Analysis Dataset

Utility	Cleaning Step	Customers		Observations	
		Treatment	Control	Treatment	Control
PG	Raw counts	151,200	20,999	5,246,130	729,162
	Subset to pre/post periods	151,999	20,999	3,509,510	487,948
	Exclude obs outside GPY4	151,200	20,999	3,509,510	487,948
	Exclude bills w/ zero days	151,200	20,999	3,509,510	487,948
	Exclude duplicate bills	151,200	20,999	3,509,510	487,948
	Exclude negative usage	151,200	20,999	3,509,510	487,948
	Exclude long/short bills	151,189	20,999	3,486,641	484,748
	Exclude outliers	151,189	20,999	3,486,253	484,701
NSG	Raw counts	91,349	21,000	3,176,556	729,603
	Subset to pre/post periods	91,349	21,000	2,122,971	487,740
	Exclude obs outside GPY4	91,349	21,000	2,122,971	487,740
	Exclude bills w/ zero days	91,349	21,000	2,122,971	487,740
	Exclude duplicate bills	91,349	21,000	2,122,971	487,740
	Exclude negative usage	91,349	21,000	2,122,971	487,740
	Exclude long/short bills	91,338	20,997	2,110,078	484,945
	Exclude outliers	91,338	20,997	2,103,250	483,272

Source: Navigant analysis of PG and NSG program tracking and customer billing data.

2.4 Statistical Models Used in the Impact Evaluation

Navigant estimated program impacts using two approaches: a simple post-program regression (PPR) analysis with lagged individual controls and a linear fixed-effects regression (LFER) analysis, both applied to monthly billing data. Navigant used the PPR results for reporting total program savings for GPY4. Both approaches should, in principal, produce unbiased estimates of program savings under a wide range of conditions, but we prefer the PPR results for two reasons. The first is that we believe, based on our own past experience analyzing the impacts of programs similar to the PG and NSG HER programs, as well as recent findings from the academic literature[12], that the savings estimates produced by the PPR model tend to be more accurate and more precisely estimated than those from the LFER model.[13] A second reason is that the implementer also uses a post-only model for their evaluation. We ran both models as a robustness check. Although the two models are structurally very different,

[12] Allcott, Hunt and Todd Rogers, 2014. "The Short-Run and Long-Run Effects of Behavioral Intervention: Experimental Evidence from Energy Conservation. *American Economic Review*, 104(10): 3003-37.

[13] One likely reason for this is that the PPR model embodies more flexibility than the LFER model, in that the former allows the individual customer control variable to vary seasonally while the latter does not – a particularly attractive feature given the highly seasonal nature of natural gas usage. The LFER model treats all unobserved inter-household heterogeneity affecting households' energy usage as time-invariant, while the PPR model uses lagged individual controls that can vary over time. This is discussed in more detail in section 6.2.1 of the Appendix.

assuming the RCT is well balanced with respect to the drivers of energy use, in a single sample we would expect them to generate similar estimates of program savings.

The LFER model combines both cross-sectional and time-series data in a single panel dataset. The regression essentially compares pre- and post-program billing data for participants and the control group to identify the effect of the program on usage. The customer-specific fixed effect is a key feature of the LFER analysis and captures all customer-specific factors affecting natural gas usage that do not change over time, including those that are unobservable. Examples of the latter include the construction and square footage of the premise, the number of occupants, the amount of seasonal sun exposure, and the thermostat settings. The fixed effect represents an attempt to control for any small, systematic differences between the treatment and control customers that might occur due to chance.

Like the LFER model, the PPR model also combines cross-sectional and time-series data in a panel dataset. Unlike the LFER model, however, it uses only the post-program data for estimation and includes the customer's lagged energy usage for the same calendar month of the pre-program period to serve as the control for any small, systematic differences between the treatment and control customers, in that sense serving the same purpose as the customer fixed effect included in the LFER model.

Section 6.1.1 of the Appendix presents the PPR and LFER models used in the analysis.

2.5 Accounting for Uplift in Other Energy Efficiency Programs

The home energy reports sent to participating households include energy-saving tips, some of which encourage participants to enroll in other PG-NSG EE programs. If participation rates in other EE programs are the same for HER participant and control groups, the savings estimates from the regression analysis are already "net" of savings from the other programs, as this indicates the HER Program had no net effect on participation in the other EE programs. However, if the receipt of HERs affects participation rates of recipients in other EE programs, then savings across all programs are lower than indicated by the simple summation of savings in the HER and the other EE programs. For instance, if the HER Program increases participation in another EE program, the resulting increase in savings may be allocated to either the HER Program or the EE program, but cannot be allocated to both programs simultaneously.[14]

As data permitted, Navigant used a difference-in-difference (DID) statistic to estimate uplift in other EE programs. To calculate the DID statistic, Navigant subtracted the change in the participation rate in another EE program between GPY4 and the pre-program year for the control group from the same change for the treatment group. For instance, if the rate of participation in an EE program during GPY4 is five percent for the treatment group and three percent for the control group, and the rate of participation during the year before the start of the HER Program is two percent for the treatment group and one percent for the control group, then the rate of uplift due to the HER Program is one percent, as reflected the following calculation:

(GPY4 treatment group participation – pre-PY treatment group participation) – (GPY4 control group participation – pre-PY control group participation) = DID statistic

(5% – 2%) – (3% – 1%) = 1%

[14] It is not possible to avoid double-counting of savings generated by programs for which tracking data are not available, such as upstream lighting programs.

The DID statistic generates an unbiased estimate of uplift when the baseline average rate of participation is the same for the treatment and control groups, or when they are different due only to differences between the two groups in time-invariant factors, such as the square footage of the residence.

An alternative statistic that generates an unbiased estimate of uplift when the baseline average rate of participation in the EE program is the same for the treatment and control groups is a simple difference in participation rates during GPY4. Navigant uses this alternative statistic –the "post-only difference" (POD) statistic –in cases where the EE program did not exist for the entire pre-program year.

Navigant examined the uplift associated with three other PG-NSG EE programs: the Home Energy Jumpstart, Home Energy Rebate, and Multifamily Energy Savings Programs. For each EE program, uplift savings were calculated separately for each of the utilities. In addition, legacy uplift (uplift from GPY3) were also calculated. These calculations are described in greater detail in section 3.2.

2.6 Process Evaluation

Navigant's GPY4 PG and NSG HER process evaluations were limited to interviews with the program implementer to update our information about the program, including plans for an additional wave of participants in GPY5. No participant surveys or interviews were pursued.

3 Gross Impact Evaluation

As shown in Table 3-1, the PG HER Program reported ex ante savings of 3,009,588 therms for GPY4. Verified savings prior to uplift adjustment was 3,318,421 therms, resulting in a verified realization rate of 1.10. Of that amount, 37,981 therms was due to uplift in other PG EE programs (17,760 therms in GPY4, and 20,221 therms of legacy uplift from GPY3). After subtracting the total uplift savings, the final verified savings was 3,280,440 therms.

Table 3-1. GPY4 PG Total HER Program Savings

Savings Category	Energy Savings (therms)
Ex Ante Savings	3,009,588
Verified Savings Prior to Uplift Adjustment *(Standard Error)*	3,318,421 *(369,798)*
Average Savings Rate (percent)	1.19
Verified Gross Realization Rate	1.10
GPY4 Uplift Adjustment	17,760
Legacy Uplift	20,221
Total Uplift Adjustment	37,981
Verified Savings After Uplift Adjustment	3,280,440

Source: Navigant analysis of PG program tracking and customer billing data.

As shown in Table 3-2, the NSG HER Program reported ex ante savings of 874,691 therms for GPY4. Verified savings prior to uplift adjustment was 1,094,406 therms, resulting in a verified realization rate of 1.25. Of that amount, -14,159 therms was due to uplift in other EE programs (-30,632 therms in GPY4, 16,473 therms of legacy uplift from GPY3). After subtracting the total uplift savings, the final verified savings was 1,108,565 therms.[15]

[15] The negative uplift resulted from recipient customers reducing their participation in one NSG EE program in GPY4 (relative to the pre-program year) at a faster rate than the control customers. This lowers the baseline and would underestimate HER program savings without the adjustment. See section 3.2 for details.

Table 3-2. GPY4 NSG Total HER Program Savings

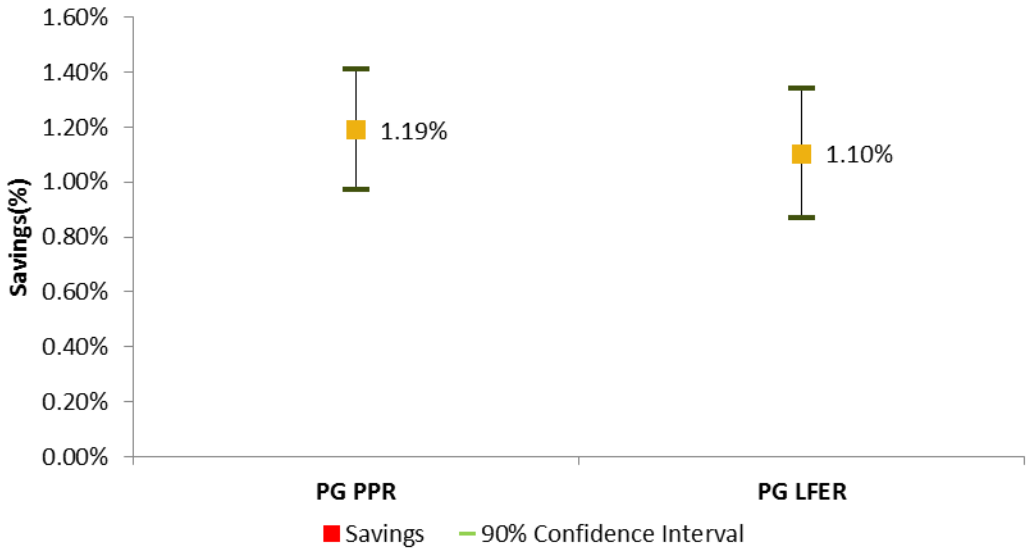
Savings Category	Energy Savings (therms)
Ex Ante Savings	874,691
Verified Savings Prior to Uplift Adjustment *(Standard Error)*	1,094,406 *(173,669)*
Average Savings Rate	0.88%
Verified Gross Realization Rate	1.25
GPY4 Uplift Adjustment	-30,632
Legacy Uplift	16,473
Total Uplift Adjustment	-14,159
Verified Savings After Uplift Adjustment	1,108,565

Source: Navigant analysis of NSG program tracking and customer billing data.

3.1 PPR and LFER Model Parameter Estimates

The PPR and LFER models generate somewhat different results for the PG and NSG HER program savings estimates in GPY4, although they do not differ significantly from each other at the 90 percent level of confidence for either program. Figure 3-1 shows the savings estimates for the GYP4 PG HER Program produced by the PPR and LFER models, bounded by their respective upper and lower 90 percent confidence intervals.

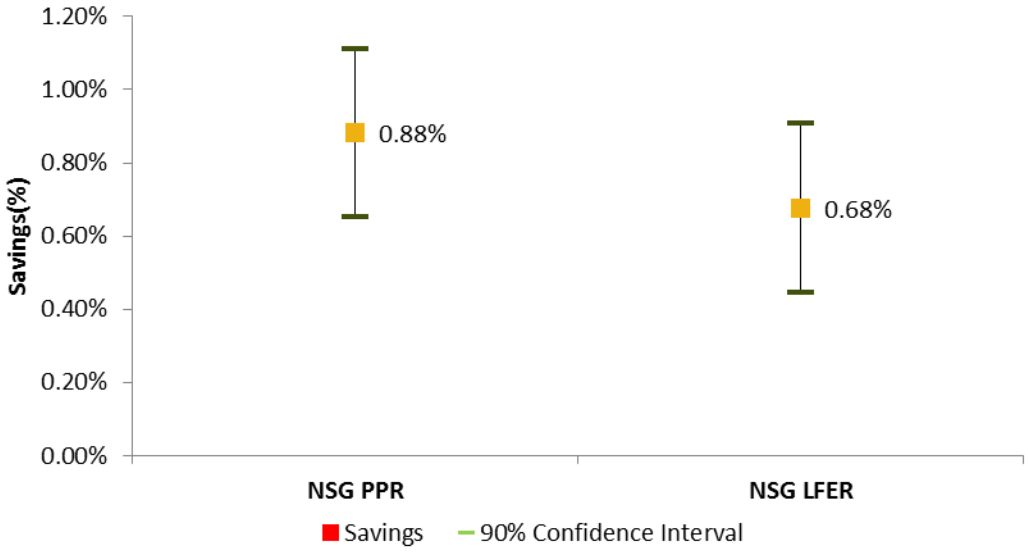
Figure 3-1. Comparison of PPR and LFER Estimates of PG GPY4 HER Savings



Source: Navigant analysis of PG customer billing data.

Note that while the savings estimates produced by the two models differ, the 90 percent confidence intervals of the PPR and LFER savings estimates overlap, indicating that the two estimates are not distinguishable in a statistically meaningful way. A similar result holds true for the NSG program, as shown in Figure 3-2.

Figure 3-2. Comparison of PPR and LFER Estimates of NSG GPY4 HER Savings



Source: Navigant analysis of NSG customer billing data.

Navigant prefers to use the PPR results for reporting total program savings for GPY4. In the past we have reported the LFER results, and both approaches should produce unbiased estimates of program savings under a broad range of conditions. However, we have shifted to reporting the PPR results for two reasons. The first is that we have found, based on our own past experience analyzing the impacts of programs similar to the PG and NSG HER programs, as well as recent findings reported in the academic literature[16], that the savings estimates produced by the PPR model tend to be somewhat more precise than those from the LFER model. One likely reason for the greater precision is that the PPR model employs a more flexible method of controlling for unobserved individual customer effects than the LFER model. The LFER approach models all sources of inter-household heterogeneity affecting customer energy usage other than those included explicitly in the model, including many that are unobservable to the evaluation team, as time-invariant, while the PPR model uses lagged own-usage as the individual control, which varies from month to month as well across customers. The benefits of this greater flexibility are discussed in more detail in Section 6.1.3 of the Appendix. A second reason is that the implementer also uses a post-only model for their evaluation. This means that comparison of our results with those produced internally by Opower provides an additional robustness check, since the main reasons for differing results should be different decisions made with respect to data cleaning. We ran both the PPR and LFER models as an internal robustness check. Although the two models are structurally

[16] Allcott, Hunt and Todd Rogers, 2014. "The Short-Run and Long-Run Effects of Behavioral Intervention: Experimental Evidence from Energy Conservation. *American Economic Review*, 104(10): 3003-37.

very different, assuming the RCT is well balanced with respect to the drivers of energy use, in a single sample we would expect them to generate similar estimates of program savings.

3.2 *Uplift of Savings in Other Energy Efficiency Programs*

The PPR estimates of program savings include savings that resulted from the uplift in participation in other EE programs caused by the HER programs. To avoid double-counting savings, program savings due to this uplift must be counted towards either the HER Program or the other EE programs, but not both programs. For the PG HER Program Navigant found 37,981 therms of total uplift savings (17,760 therms in GPY4, and an additional 20,221 therms of legacy uplift from GPY3), which amounts to roughly 1.3 percent of total verified savings prior to uplift adjustment. Subtracting these savings from the PG HER program's verified gross savings (3,159,964 therms) yields a final verified net savings estimate of 3,121,983 therms for GPY4. The details of the PG HER uplift calculations are shown in Table 3-3.

Table 3-3. GPY4 PG HER Uplift Adjustment Details

Category	Program		
	Home Energy Jumpstart	Home Energy Rebate	MESP
Median program savings (annual therms per participant)*	38	139	102
# HER treatment households	151,200	151,200	151,200
Rate of participation, PY2 (%)	1.65%	0.44%	0.02%
Change in rate of participation from pre-program year (%)	0.36%	-0.67%	-0.01%
# HER control households	21,000	21,000	21,000
Rate of participation, PY2 (%)	1.54%	0.43%	0.02%
Change in rate of participation from pre-program year (%)	0.29%	-0.73%	-0.02%
DID statistic	0.07%	0.06%	0.01%
Change in program participation due to HER Program	106	86	18
Statistically significant at the 90% confidence level?	No	No	Yes
Savings attributable to other programs (therms)	**4,046**	**11,898**	**1,815**
Legacy Uplift	17,076	3,145	0
Total Uplift	**21,122**	**15,043**	**1,815**

Source: Navigant analysis of PG program tracking and customer billing data.

* Median program savings are the median therms impacts of HER recipients in each program.

For the NSG HER Program Navigant found -14,159 therms of uplift (-30,632 therms in GPY4 and 16,473 therms of legacy uplift from GPY3). The negative uplift in GPY4 was caused by HER Program participants reducing their participation in the NSG Home Energy Rebate program in GPY4 (relative to their participation in the pre-program year) at a faster rate than the corresponding group of control customers. This phenomenon lowers the baseline against which the HER program savings is measured, and would lead to an underestimate of the HER program's savings without the adjustment. Subtracting the total uplift from the NSG program's verified gross savings (1,115,401 therms) raises the HER's verified net savings to 1,129,559 therms. The details of the NSG HER uplift calculations are shown in Table 3-4.

Table 3-4. GPY4 NSG HER Uplift Adjustment Details

Category	Program		
	Home Energy Jumpstart	Home Energy Rebate	MESP
Median program savings (annual therms per participant)*	40	135	73
# HER treatment households	91,350	91,350	91,350
Rate of participation, PY2 (%)	0.79%	0.74%	0.02%
Change in rate of participation from pre-program year (%)	0.53%	-0.75%	-0.17%
# HER control households	21,000	21,000	21,000
Rate of participation, PY2 (%)	0.74%	0.84%	0.02%
Change in rate of participation from pre-program year (%)	0.47%	-0.48%	-0.18%
DID statistic	0.06%	-0.27%	0.01%
Change in program participation due to HER Program	55	-249	10
Statistically significant at the 90% confidence level?	No	Yes	No
Savings attributable to other programs (therms)	**2,207**	**-33,568**	**729**
Legacy Uplift	9,547	6,926	0
Total Uplift	**11,755**	**-26,642**	**729**

Source: Navigant analysis of NSG program tracking and customer billing data.

* Median program savings are the median therms impacts of HER recipients in each program.

4 Net Impact Evaluation

A key feature of the design of the PG and NSG HER programs is that the statistical analyses used to generate the energy savings estimates inherently yield net savings because of the RCT program design. Since both are opt-out programs, there is no possibility of there being any participants who might have opted to receive the individualized home energy reports in the absence of the program. While it is possible that some customers receiving the reports might have undertaken the energy-conserving actions or high-efficiency equipment purchases that they did in GPY4 anyway, the random selection of program participants and controls ensures that the participant and control groups of customers will exhibit the same degree of energy-conserving behavior and purchases. Thus, the expected value of free-ridership is zero, and no "net-to-gross" (NTG) adjustment is necessary.

5 Findings and Recommendations

This section summarizes the key impact and process findings and recommendations. …

Verified Net Savings, Gross Savings and Realization Rate.

Finding 1. Overall, both programs improved their savings performance in GPY4 relative to GPY3. In GPY4 the PG HER Program achieved verified net savings of 3,318,421 therms prior to uplift adjustment, which represents a savings rate of 1.19 percent. In GPY3, the corresponding savings was 2,072,182 therms, or 0.85 percent. In PGY4 the NSG program achieved verified savings of 1,094,406 therms prior to uplift adjustment, which represents a savings rate of 0.88 percent. In GPY3, the corresponding figure was 662,518 therms, or 0.62 percent.

Finding 2. The therms savings reported in Finding 1 are not strictly comparable across program years for two reasons:

- The PG and NSG HER programs did not start until October 2013, which means that they only ran for eight months of GPY3;
- The numbers of participants in each program in GPY4 were somewhat reduced from those prevailing at the rollout of the program in GPY3 as a result of due to recipient customer drop-outs and recipient and control customer move-outs. (Recipients may opt out of the program at any time, and move-outs occur at a low but relatively constant rate in both groups. Neither type of attrition is replaced.)

Thus, the more apt comparison between program years for the purpose of judging relative performance is on the basis of the percent savings rates the programs achieved in each year. The savings rates improved year-over-year for both programs: from 0.85 percent to 1.19 percent for the PG program, a 0.34 percentage point increase (an improvement of 40 percent over GPY3); and from 0.62 percent to 0.88 percent, a 0.26 percentage point increase (an improvement of 42 percent over GPY3) for the NSG program.

Finding 3. Opower indicated in a discussion with the evaluation team in 2015[17] that they intended to roll out a second wave of approximately 12,000 new participants in GPY5 in the NSG service territory, along with approximately 10,000 new controls. They said they expected reports to start being sent out to the new wave of participants before the end of 2015.

Finding 4. With the addition of 12,000 new HER recipients and 10,000 new control customers in the NSG service territory in GPY5, Navigant suspects that NSG and Opower may have tapped out the potential for expanding the program. We understand that given the small size of the NSG pool of qualifying residential customers, Opower's ability to target the program to customers was constrained even prior to the advent of the new wave in GPY5.[18]

Recommendation 1. In view of the improved performance of the PG and NSG HER programs in GPY4 – in particular the NSG program's 1.25 realization rate – as well as the advent of a new wave of recipients in the NSG service territory in GPY5, Navigant recommends that both utilities consider undertaking persistence research on their HER programs beginning in GPY6. This research, which would entail randomly selecting a subset of recipient customers

[17] Telephone interview with Opower project managers, May 19, 2015.
[18] *Ibid.*

who would have their reports terminated, would permit measurement of the rate at which program savings rates decay once reports cease, and thus provide insight into persistence and measure life.

6 Appendix

6.1 Detailed Impact Methodology

Navigant used two regression models to estimate impacts: a PPR model and an LFER model. The following sections present each model.

6.1.1 PPR Model

The PPR model controls for non-treatment differences in natural gas use between treatment (i.e., home energy report recipient) and control (non-recipient) customers using the customer's own lagged energy use as an explanatory variable. In particular, the model frames therms used in calendar month *t* of the post-program period as a function of both the treatment variable and therms used in the same calendar month of the pre-program period. The underlying logic is that systematic differences between control and treatment customers will be reflected in differences in their past energy use, which is highly correlated with their current energy use. Since lagged customer usage varies monthly, it is able to capture systematic differences between treatment and control customers that vary seasonally. Formally, the model is the following:

Equation 1: $$ADU_{kt} = b_1 Treatment_k + b_2 ADUlag_{kt} + \sum_J b_{3j} Month_{jt} + \sum_J b_{4j} Month_{jt} \times ADUlag_{kt} + e_{kt}$$

where

ADU_{kt} is average daily therms consumption by household *k* in bill period *t*

$Treatment_k$ is a binary variable taking a value of 0 if household *k* is assigned to the control group, and 1 if assigned to the treatment group

$ADUlag_{kt}$ is household *k*'s average daily therms usage in the same calendar month of the pre-program year as the calendar month of month *t*

$Month_{jt}$ is a binary variable taking a value of 1 when j = t and 0 otherwise[19]

e_{kt} is the cluster-robust error term for household *k* during billing cycle *t;* cluster-robust errors account for heteroskedasticity and autocorrelation at the household level.[20]

The coefficient b_1 is the estimate of average daily therms savings due to the program in PY6.

[19] In other words, if there are *T* post-program months, there are *T* monthly dummy variables in the model, with the dummy variable $Month_{tt}$ the only one to take a value of 1 at time *t*. In short, they are monthly fixed effects designed to capture the net effect on gas usage of all month-specific effects affecting usage that are common across all households, for example weather effects.

[20] Ordinary least-squares (OLS) regression models assume that the data are homoskedastic and not autocorrelated. 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. If either of these assumptions is violated, the resulting standard errors of the parameter estimates will be biased (usually underestimated).

6.1.2 LFER Model

The simplest version of an LFER model convenient for exposition is one in which average daily consumption of natural gas by household *k* in bill period *t*, denoted by ADU_{kt}, is a function of the following three terms:

1. The binary variable $Treatment_k$
2. The binary variable $Post_t$, taking a value of 0 if month *t* is in the pre-treatment period, and 1 if in the post-treatment period
3. The interaction between (product of) these variables, $Treatment_k \cdot Post_t$

Formally, as shown in the following equation:

Equation 2: $$ADU_{kt} = a_{0k} + a_1 Post_t + a_2 Treatment_k \times Post_t + e_{kt}$$

In this model, the coefficient a_{0k} captures all household-specific effects on natural gas consumption that do not change over time, including those that are unobservable to the evaluation team. The coefficient a_1 captures the average effect *across all households* of being in the post-treatment period, while the coefficient a_2 captures effect of being *both* in the treatment group *and* in the post period—in other words, the effect directly attributable to the program. Put another way, whereas the coefficient a_1 captures the change in average daily usage between the pre- and post-treatment for the *control* group, the sum $a_1 + a_2$ captures this change for the treatment group, and therefore the difference, a_2, estimates the average daily therms savings due to the program in GPY4.

6.1.3 Detailed Impact Results: Parameter Estimates

Table 6-1 presents, for each utility program, two sets of regression results: one for each of the models described in the previous section. Specifically, it presents each model's estimate of the average daily therms savings due to the HER Program in GPY4, and its standard error. For the PPR model, presented in section 6.1.1, this is the coefficient b_1 in Equation 1. For the LFER model, presented in section 6.1.2, it is the coefficient a_2 in Equation 2.

Table 6-1. Savings Parameter Estimates

Utility	PPR Model		LFER Model	
	Parameter Estimate	*Standard Error**	Parameter Estimate	*Standard Error**
PG	-0.06205	*0.00692*	-0.05737	*0.00744*
NSG	-0.03403	*0.00540*	-0.02632	*0.00548*

Source: Navigant analysis of PG and NSG program tracking and customer billing data.
* Cluster-robust standard errors.

As mentioned in Section 3.1 above, both the PPR and LFER models are expected to produce unbiased estimates of program savings under a broad range of conditions, but Navigant believes that on balance, the PPR model is the preferred approach. To understand why, it helps to think consider how the two models differ. Fundamentally, the difference between the two models boils down to how each one attempts to handle "inter-customer heterogeneity." By this we mean all of the myriad ways in which individual customers differ in terms of the characteristics that affect how much natural gas they consume in a given month. Residential customers' average usages vary for a lot of reasons besides whether or not they are receiving home energy reports. Obvious examples include variations in the size, construction, orientation, and degree of shading of the premise; the number and ages of the occupants of the home; the types, sizes and vintages of heating and air conditioning systems, water heating systems, and other appliances; the amount and quality of the home insulation; and so on.

Generally speaking, these are all important determinants of energy usage, and none of them is typically known to the evaluation team, which leaves the problem of how best to accommodate their effects in our statistical model to get the best possible estimate of program savings. Referring back to Equation 1, we see that the PPR model says that a customer's average daily usage of natural gas is a function of whether they are receiving home energy reports or not, their average usage in the same month during the pre-program year, what month of the year it currently is, plus an error term, which is the repository that holds all of the current month's usage that the rest of the model can't explain. The idea behind the PPR model is that, whatever the combination of unobserved customer characteristics may be for any particular customer, they're likely to have been very similar in the pre-program year, and so by including the lagged term in the model, their net combined effect on current usage will be differenced out month by month. Referring to Equation 2, the LFER model instead uses an individual intercept to do the same thing as the lagged usage term.

Thinking of these unobserved differences among customers as differences in their stock of structures and equipment, it seems reasonable to treat them as time-invariant for any given customer.[21] However, while the stocks themselves may not change over time, their impacts on household energy usage are *not* time-invariant. A fixed-effects model essentially gives each household in the sample a separate, customer-specific parameter that shifts their entire annual natural gas consumption function up or down from the overall mean by a fixed amount. The problem comes when a particular household tends to consume *more* gas than average during part of the year, but *less* than average during the rest of the year. Consider the example of one household that heats with a central gas-fired furnace in the winter but has an electric water heater, and another household that heats in the winter with baseboard electric heat but has a natural gas water heater. The former may use more gas than average during the winter heating season but less than average in the spring, summer and fall, while the reverse is likely true of the latter. In both cases, the LFER model attempts to summarize the net effect of all of the customer's unobserved characteristics that affect each customer's gas usage with a single parameter, but is faced with the need to have those systematic effects vary by season of the year. Since the LFER model isn't able to accommodate seasonally-varying individual effects, it ends up averaging the two over the full year, which means that

[21] In fact, households *do* change their appliances and home mechanicals occasionally, and household demographics do change slowly over time as children are born, grow up, and leave the nest. But for the purposes of evaluating an energy efficiency program, these changes assumed to be sufficiently rare that they can be safely ignored.

more of the customer's variations in energy usage from month to month will be left unexplained by the model and instead relegated to the error term. The PPR model, which is able to accommodate systematic differences in the individual customer effect, doesn't face this problem. For that reason, we believe it is probably yields more accurate and precise results because of this added flexibility.