Home Performance with ENERGY STAR® Program Impact Evaluation Report (PY2010–2013)

Final Report

Volume 2: Phase 1 Billing Analysis

(PY2010–2011)

November 21, 2016

Prepared for

The New York State
Energy Research and Development Authority

Carley Murray
Project Manager

Prepared by

ERS
120 Water Street, Suite 350
North Andover, MA 01845
Phone: 978-521-2550

Principal Investigator:
West Hill Energy and Computing

NOTICE

ABSTRACT

Volume 2 presents methods and findings from Phase 1 of NYSERDA's Home Performance with ENERGY STAR® Program ("HPwES" or "Program") impact evaluation. The Phase 1 billing analysis was designed to estimate household first-year energy savings including all program year (PY) 2010/2011 participants with sufficient pre- and post-installation consumption data. The purpose of this study was to provide robust and reliable estimates of first-year energy savings, both electric and natural gas, at the household and program level.

ACKNOWLEDGMENTS

The ERS Impact Evaluation Team would like to thank Carley Murray and Judeen Byrne for the assistance they provided throughout the process of implementing the HPwES impact evaluation. The ERS Team would also like to thank NYSERDA staff, including Karen Hamilton, Laura Geel, and Kim Lenihan, who provided critical support and information to help us understand the Program procedures and policies. In addition, the ERS Team appreciates the valuable feedback provided by William Saxonis of the Department of Public Service and Rick Ridge and Ralph Prahl from the DPS contractor team, through the review of our planned work. Finally, the ERS Team would like to acknowledge the efforts of the utility staff members who provided the electric consumption information and records requested for participant projects.

TABLE OF CONTENTS

NOTICE I

ABSTRACT II

ACKNOWLEDGMENTS III

LIST OF TABLES V

LIST OF FIGURES V

SECTION 1: INTRODUCTION 1

SECTION 2: PROGRAM DESCRIPTION 2

2.1 Summary of Program-Reported Savings 2

SECTION 3: METHODS 6

3.1 Data Sources and Issues 6

3.2 Baseline 9

3.3 Preparation of Billing records 10

3.4 Billing Analysis 11

SECTION 4: RESULTS 16

4.1 Gross Energy Savings Results 16

4.2 Discussion 23

SECTION 5: CONCLUSIONS 35

SECTION 6: REFERENCES 36

LIST OF TABLES

Table 1. Total Program Reported Natural Gas Savings by End Use ... 3
Table 2. Total Program Reported Extra Natural Gas Usage ... 4
Table 3. Program Reported Electric Savings by End Use ... 4
Table 4. Program Reported Extra Electricity Usage ... 5
Table 5. Attrition in HPwES Billing Models ... 9
Table 6. Comparison of Savings for HPwES Billing Analyses ... 16
Table 7. Comparison of Electric Model Results and Program Reported Savings ... 17
Table 8. Summary Statistics from the Natural Gas Regression Model ... 19
Table 9. Informational Purposes Only- Natural Gas Program Reported and Evaluated Savings by Measure Group ... 20
Table 10. Summary Statistics from the Electric Regression Model ... 22
Table 11. Comparison of Program Reported Savings and Heating Consumption Patterns for Homes with Electric Heating Measures ... 25
Table 12. Comparison of Savings and Consumption for Cooling Measures ... 26
Table 13. Distribution of Homes with Program Reported A/C Savings by Utility ... 27
Table 14. Incidence of Heating System Replacements and Repairs ... 28
Table 15. Comparison of Heating System Measure Combinations ... 29
Table 16. Natural Gas Realization Rates by Savings Level ... 30
Table 17. Electric Realization Rates by Savings Level ... 30
Table 18. Natural Gas Savings and Realization Rates by Loan Source ... 31
Table 19. Realization Rates by Contractor ... 32
Table 20. Realization Rates by Contractors' Length of Program Participation ... 33
Table 21. Comparison of Annual Consumption and Evaluated Savings ... 34

LIST OF FIGURES

Figure 1. Program Reported Natural Gas Savings by End Use ... 3
Figure 2. Program Reported Electric Savings by End Use ... 5

SECTION 1: INTRODUCTION

Volume 2 presents the methods and findings from Phase 1 of NYSERDA's Home Performance with ENERGY STAR® Program ("HPwES" or "Program") impact evaluation. This report provides a description of the HPwES Program, outlines the impact evaluation approach and methods, and presents the results of Phase 1. This evaluation cycle covers projects completed in program years (PYs) 2010 and 2011, and the Phase 1 billing analysis is the first step of a more comprehensive impact evaluation.

Phase 1 of the current evaluation cycle was designed to estimate household first year energy savings using a billing analysis including all PY 2010/2011 program participants with sufficient pre- and post-installation consumption data. The purpose of this study was to provide robust and reliable estimates of first year energy savings, both electric and natural gas, at the household and program level.

The most recent impact evaluation of HPwES for PYs 2007 and 2008 was completed by the Megdal & Associates Impact Evaluation Team in 2012. The 2007 and 2008 evaluation was the first to utilize the participant consumption data. As a result, the Evaluation Team experienced some difficulties obtaining and interpreting the data from each of the utilities and the final billing analysis included data provided by three of the seven utilities.

Subsequently, process and data confidentiality protocols were put in place with NYSERDA and each of the utilities. These processes facilitated the delivery of consumption data from all of the electric and natural gas utilities. Phase 1 of the PY 2010/2011 impact evaluation expanded the scope of the billing analysis to include all of the utilities.

The remaining sections of this report provide a brief description of the Program, a discussion of Phase 1 methods, results of the analysis, and conclusions.

SECTION 2: PROGRAM DESCRIPTION

The Home Performance with ENERGY STAR® Program ("HPwES" or "Program") encourages home and building owners and tenants of existing one- to four-family homes to implement comprehensive energy efficiency improvements and technologies by contractors accredited by the Building Performance Institute and participating in the HPwES Program. Eligible measures include building shells, such as air sealing and insulation; appliances, such as ENERGY STAR refrigerators; heating, such as boilers and furnaces; cooling, such as ENERGY STAR room or central air conditioners; and certain renewable energy technologies.

The HPwES Program is designed to offer enhanced assistance to low- to moderate-income households. The "Assisted" component of the Program is available to residents with up to 80% of the area median income, or 80% of the state median income – whichever is higher for the county.

A wide range of measures are installed through this program, from screw-in compact fluorescent lamps (CFLs) and replacement refrigerators to insulation and heating system replacements. While the vast majority of measures have net energy savings, specific measures sometimes result in savings from one fuel and additional use of another fuel. This situation occurs due to measures that are designed to shift energy use from one fuel to another and from some measures that are primarily designed to save one fuel but may cause a small increase in consumption of another fuel. Replacing electric baseboard heating with a natural gas furnace is an example of fuel switching. Heat pumps are another example of measures that save energy overall but may result in an increase in either natural gas or electric consumption in comparison to the previous energy use in the household.

2.1 SUMMARY OF PROGRAM-REPORTED SAVINGS

Program tracking by measure is highly detailed and both savings and extra use are recorded for each fuel type. The following tables and figures show the program-reported savings for 2010 and 2011. To simplify the presentation, program reported impacts were divided into savings and extra use for electricity and for natural gas.

Table 1 and Figure 1 show the total program-reported natural gas savings by end use. Over 90% of the natural gas savings were associated with heating-related measures, such as envelope upgrades (insulation and air sealing), heating system repair, and replacement heat pumps.

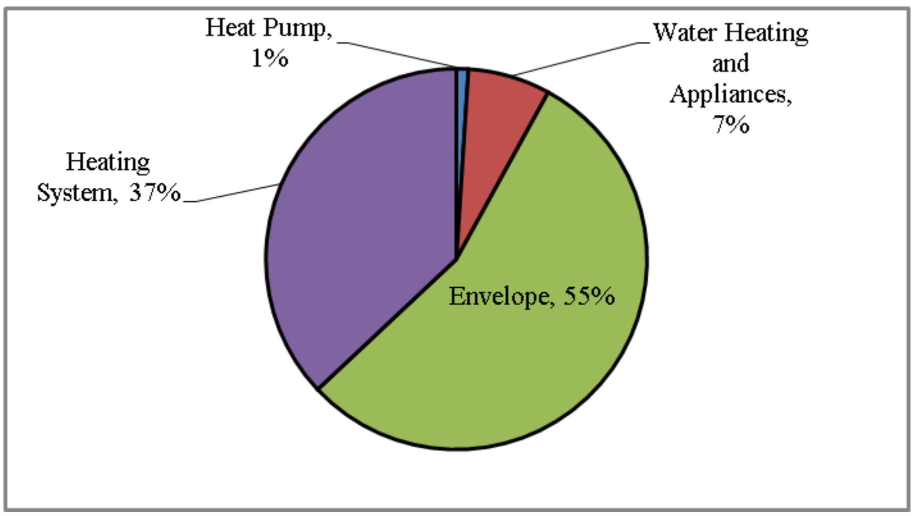
Table 1. Total Program Reported Natural Gas Savings by End Use

End Use	Number of Homes with Measure	Natural Gas Savings (MMBtu/year)	Natural Gas Savings per Home with Measure (MMBtu/year)	Percentage of Total Savings
Envelope[1]	7,468	162,939	21.8	55%
Heating system	8,194	107,846	13.2	37%
Water heating & appliances	3,542	21,517	6.1	7%
Heat pump	32	2,232	69.8	1%
Total[2]	9,917	294,534	29.7	100%

[1] Envelope measures include attic, wall, and basement insulation, air sealing, and replacement windows and doors.

[2] The rows in this column do not add to the total as some participants installed measures in multiple end uses. The total number of homes reflects the unique number of households installing any measure that generated extra use.

Figure 1. Program Reported Natural Gas Savings by End Use



The following table shows the distribution of the natural gas extra use. Most of the extra use (70%) was related to space heating fuel switches from another fuel to natural gas. The extra use associated with lighting, refrigerators, and appliances was due to the waste heat penalty, *i.e.*, the assumption that installing more efficient lighting and appliances reduces the waste heat provided by these devices and that this waste heat will need to be supplied by the heating system. While the waste heat penalty is small on a house-by-house basis, it adds up to 13% of the total extra use estimated by the program. In the previous impact evaluation, the billing models were not able to pick up this particular type of extra use, and to date, this theoretical estimate of the waste heat penalty has not been demonstrated through a billing analysis. Table 2 shows the total program-reported extra natural gas usage by end use.

Table 2. Total Program Reported Extra Natural Gas Usage

End Use	Home with Measure	Natural Gas Extra Use (MMBtu/year)	Average Extra Use per Home (MMBtu/year)	Percentage of Total Extra Use
Heating system fuel switch	353	26,174	74.1	70%
Hot water fuel switch	295	5,853	19.8	16%
Lighting[1]	2,792	4,339	1.6	12%
All other heating measures	332	513	1.5	1%
Refrigerator & other appliances[1]	942	467	0.5	1%
Total[2]	3,420	37,372	10.9	100%

[1] The extra use associated with these end uses is most likely due to the waste heat penalty.

[2] The rows in this column do not add to the total as some participants installed measures in multiple end uses. The total number of homes reflects the unique number of households installing any measure that generated extra use.

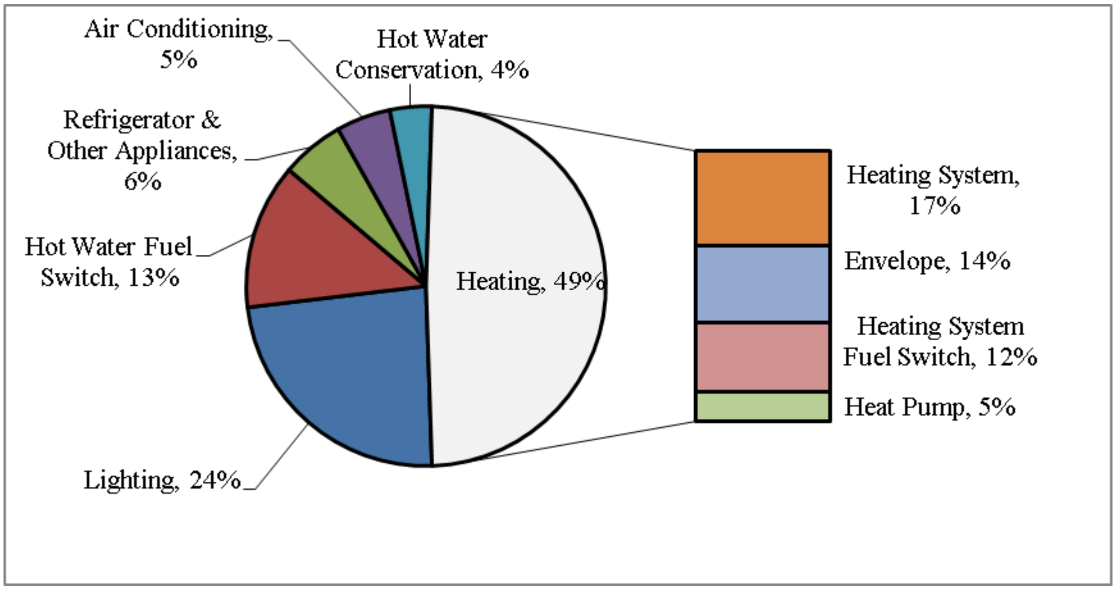
Table 3 and Figure 2 present the program reported electric savings by end use. Almost 50% of the electric savings were associated with heating related measures, with heating system measures accounting for 17%, envelope updates 14%, space heat fuel switches 12% and heat pumps 5%. Lighting accounts for the bulk of the remaining savings.

Table 3. Program Reported Electric Savings by End Use

End Use	Homes with Measure	Electric Savings (kWh/year)	Savings per Household with Measures (kWh/year)	Percentage of Total Savings
Heating-Related Measures				
Heating system	4,959	1,529,629	308	17%
Envelope	3,291	1,250,474	380	14%
Heating system fuel switch	86	1,116,062	12,977	12%
Heat pump	118	478,994	4,059	5%
Other Measures				
Lighting	3,421	2,112,960	618	24%
Hot water fuel switch	255	1,168,701	4,583	13%
Refrigerator & other appliances	1,357	508,078	374	6%
Air conditioning	1,489	437,414	294	5%
Hot water conservation	1,260	336,696	267	4%
Total[1]	8,252	8,939,008	1,083	100%

[1] The rows in this column do not add to the total as some participants installed measures in multiple end uses. The total number of homes reflects the unique number of households installing any measure that generated extra use.

Figure 2. Program Reported Electric Savings by End Use



Extra use was also associated with some electric measures, as shown in Table 4. This analysis shows that the vast majority of the electric extra use is from heat pumps.

Table 4. Program Reported Extra Electricity Usage

End Use	Number of Homes with Measure	Electric Extra Use (kWh/year)	Average Extra Use per Home (kWh/year)	Percentage of Total Extra Use
Heat pump	185	1,227,466	6,635	84%
Water heating fuel switch	40	151,129	3,778	10%
Heating measures	1,930	66,584	34	5%
Other	218	12,092	55	1%
Total[1]	2,222	1,457,271	656	100%

[1] The rows in this column do not add to the total as some participants installed measures in multiple end uses. The total number of homes reflects the unique number of households installing any measure that generated extra use.

SECTION 3: METHODS

This section describes the methods used to develop the realization rates (RRs). Evaluated savings were estimated from a billing analysis. The remainder of this section covers the data sources and issues, attrition, data preparation, and billing analysis.

3.1 DATA SOURCES AND ISSUES

A billing analysis requires three major types of data:

- The program data on measures installed in each home and characteristics of the homes
- The consumption history pre- and post-retrofit (billing records) from the electric and gas utilities
- Weather data from the National Oceanic and Atmospheric Administration (NOAA)

The program data was used to identify the major measure groups installed in each home and the project completion date. The heating and cooling degree days for each billing cycle were calculated from the NOAA data and connected to the billing data through the weather station.

3.1.1 Data Cleaning

Data cleaning is a critical component of any billing analysis and is generally the most time-consuming step in the process. The Impact Evaluation Team carefully reviewed the billing data for the following issues:

- Sufficient period of pre- and post-installation billing records; nine months pre-installation and nine months after the installation of all measures is desired, with at least three consecutive winter months for the natural gas model in both the pre- and post-installation periods.
- Breaks in billing history indicating a possible lapse in service – typically monthly reads with no energy use or missing reads
- Overall use is within the range of residential use; homes with electric space heat tend to have higher use; above 50 MWh per year suggests some other type of use rather than residential; sometimes the consumption level is lower than would be used to run a refrigerator and a few lights, suggesting that the home may be unoccupied for some periods.
- Consumption with high variability or a “see-saw” pattern, which may indicate unmarked estimated reads

Homes identified through this process were not necessarily eliminated from the model, but they were identified and reviewed for inclusion.

3.1.2 Attrition in the Billing Models

Bias and sampling precision are two critical factors that affect the underlying reliability of evaluation results. There is no sampling for a large-scale regression model, as was conducted for the Home Performance with ENERGY STAR® Program ("HPwES" or "Program"), as all participants with sufficient billing history were included in the models. Thus, the primary concern for this evaluation was the possibility of bias.

Two primary types of bias were considered. The first potential source of bias is that participants without available or sufficient consumption history were excluded from the model (attrition). The second is the extent to which external influences could create change in energy usage and affect the results of the billing analyses.

The concern regarding attrition is whether the removal of specific groups of homes with similar characteristics may introduce bias into the regression results. The potential impacts of attrition are dependent on the relationship between the type of homes removed from the model and services offered by the program, as well as the methods used to conduct the analysis and calculate evaluated program savings.

Two aspects of the selected evaluation methods were designed to minimize the impacts of attrition, as explained below.

- **Analysis Method:** RRs were determined by measure group for homes in the model and then applied by measure group to the program activity, which assumes that the RRs are similar between the homes in the model and the total program. The alternative approach of estimating the evaluated gross savings per measure group would assume that the estimated savings per home (which would reflect the size of the homes and the climate zones of the homes in the model) are the same between the model and the total program projects.
- **Regression Model:** A fixed-effects regression model was used. The fixed-effects model compares each home to itself, which means that house-specific differences that are consistent across the analysis period are addressed in the regression analysis.

Thus, in assessing the potential bias associated with attrition from the billing analysis, the key issue is whether there is any expectation that specific groups of homes have different RRs rather than whether the homes in the model are a good match to the homes in the population.

Some of the critical factors that are unlikely to be affected by attrition in the context of this study and the applied methods are weather effects (directly included in the model), the fixed characteristics of the homes (housing stock, appliance holdings, etc.), the mix of measures, and program delivery strategies (which were constant over all projects). Issues that were considered as potentially introducing bias into the regression results are described below.

- If specific large contractors are better or worse at estimating household savings and if all homes completed by one or more of these large contractors are completely removed from the analysis, attrition could present a potential source of bias. Some participants moved during the analysis period and cannot be included in the analysis since the billing records do not cover the critical months before and after the installation. Thus, participants who tend to move often are effectively removed from the analysis as a group. Although there is no way to assess the impact of this effect or modify the billing models to address it, this issue is more likely to be prevalent in the low-income sector and is unlikely to introduce bias to the HPwES models.

Potential bias from these sources on the RR could be either upward or downward.

Table 5, below, shows the attrition in current billing analysis for the PY 2010–2011; for comparison purposes, the table also includes the attrition from the two previous billing analyses. For natural gas, the PY 2010–2011 analysis included seven utilities; billing data was provided for 70% of the participating homes with savings and almost half of all participants with savings were included in the billing analysis.

No billing data was obtained for about 30% of the participants with natural gas savings, most likely due to issues with matching the NYSERDA Program records to the utility billing records. This lack of billing data accounts for about 60% of the overall attrition, and participating homes eliminated for this reason are expected to be randomly distributed and would not introduce bias into the results, *i.e.*, there is no reason to connect the magnitude of savings in these homes to the ability to obtain the billing records. The remaining 40% of the attrition is due to participants with insufficient billing records, which is likely to be largely due to changes in occupancy and would also not be expected to introduce bias.

In contrast, the PY 2007–2008 natural gas billing model included only two utilities and 10% of all participating homes with savings.[1] Thus, the attrition for the PY 2010–2011 billing models was a substantial improvement over the previous evaluation. This level of attrition is consistent

[1] The high attrition in the PY 2007–2008 impact evaluation was mostly due to the fact that this evaluation was the first time NYSERDA requested billing data from the utilities and it took quite a while to work out the issues with the data collection process.

with other, similar billing analyses. Table 5 summarizes the attrition in the HPwES billing models.

Table 5. Attrition in HPwES Billing Models

	2010-2011		2007-2008	
	Electric Service	**Natural Gas Service**	**Electric Service**	**Natural Gas Service**
Number of projects	13,811	12,642	15,350	15,130
Number of homes[1]	12,581	11,453	14,925	14,710
Number of homes with savings	8,571	10,256	8,558	14,187
Number of homes with savings and billing data	6,196	7,128	7,794	5,054
Total participants in analysis after cleaning[2]	3,185	5,009	2,536	1,462
Percentage of homes with billing data included in analysis	51%	70%	33%	29%
Percentage of homes with savings included in analysis	37%	49%	30%	10%
Total number of utilities included in analysis	7	7	3	3

[1] More than one project was completed in some homes. Consequently, the number of homes is slightly smaller than the number of projects.
[2]The large number of participants/households with billing data eliminated from the 2007–2008 models reflects the fact that billing data from NYSEG and RG&E contained a large number of unidentified estimated reads and reconciliations. The final models were run both with and without data representing these two utilities. The statistical reliability of the analysis dropped dramatically and the estimated savings from the model were substantially lower when these utilities were included in the model, thus they were ultimately excluded.

3.2 BASELINE

Since HPwES is primarily a retrofit program, the baseline is the pre-installation condition. In the billing analysis, the pre-installation conditions are reflected in the pre-installation billing records. An exception to the retrofit scenario occurs when participants replace an aging or nonfunctional heating or cooling system. In this case, the decision is not to remove a piece of working equipment in order to achieve energy savings, but rather to replace equipment that has failed or is expected to fail in the near future. The baseline for these natural (market opportunity) replacements should be the state or federal standard rather than the consumption of the previously-existing equipment.

The incidence of replacement on failure was investigated in the previous evaluation, with uncertain results; further research will be conducted as part of Phase 1 of the HPwES PY2010/2011 Impact Evaluation. This Phase 1 billing analysis was conducted assuming that the pre-existing conditions are the appropriate baseline.

3.3 PREPARATION OF BILLING RECORDS

A step in the data cleaning and preparation for this study was “calendarization,” which refers to the allocation of the energy in the individual bills to actual annual calendar months. Calendarization allows every read to be placed on equal footing, e.g., the calendar month. This avoids a situation in which a month with a truncated read cycle with few days has the same impact on the model results as a full or long read cycle. The potential disadvantage is that it requires estimating the consumption that occurred during the confines of the calendar month. Thus, both strategies; calendarization and using the billing data as is, have potential advantages and disadvantages. For this analysis, the billing data was calendarized.

The calendarization process begins by merging the billing and the weather data by read dates to ensure that the weather and billing data are appropriately aligned. While it depends on the length of the billing interval, a single bill typically includes days in more than a single calendar month. Using the previous read date and the read date for each billing cycle, the number of billing days and total degree days can be allocated to the different calendar months associated with a billing interval. In the case of a 28-day bill from December 10th, for example, ten of the days (and the respective degree days) are associated with December and the remaining 18 are associated with November.

The counts of days or sum of degree days are then used to calculate shares that will allocate the billing data consumption to each of the calendar months. For natural gas, which is strongly associated with weather, the allocation of consumption during heating months relies on the share of degree days. For electric consumption and the nonheating months in the natural gas model, the shares are based solely on the number of days.

This process is conducted for every billing observation and then the allocated consumption is summed to the calendar month. The final step is to normalize the billing consumption to 30.4 days per month to provide each month with the same length for the billing analysis.

3.4 BILLING ANALYSIS

Billing analysis was selected for estimating gross savings in this impact evaluation due to the characteristics of the HPwES program. Billing analysis is an effective tool for impact evaluation for retrofit programs when savings are estimated from the existing condition in the home prior to the installation and the savings are of sufficient magnitude to be found in the billing records. For models with only a few homes, the general rule of thumb is that the savings should be about 10% of the overall consumption. Models with more homes can be used to estimate smaller savings.[2]

The billing analysis was conducted using a cross-sectional, time-series regression to estimate program savings. A fixed effects model was used to address the energy-related characteristics of the home that do not change over time, such as the size of the home, and the presence of major electric or natural gas appliances and heating equipment. The regression model included weather and efficiency installations through the program as predictor (independent) variables. The response (dependent) variable was the monthly energy consumption, and the regression coefficients for program variables were used to estimate the program savings. Separate natural gas and electric models were developed.

The model was a generalized linear model with customer-specific intercept of the form shown in the equation below.

$$C_{it} = \alpha_i + \tau_t + \sum_{j=1}^{p} x_{ijt}\beta_j + \sum_{k=1}^{q} z_{ikt}\gamma_k + \varepsilon_{it} \qquad (1)$$

where,

C_{it} is the monthly consumption for the household *i* in period *t*, expressed in monthly kWh per day,

α_i is the "customer-specific" intercept (or error) for household *i*, accounting for unexplained difference in use between households associated with the number of occupants, appliance holdings and lifestyle,

[2] Billing analysis was employed in the most recent impact evaluation and lessons learned from this study will be applied to improve future study efforts. In this evaluation, there were over 3,000 homes in the electric model and 5,000 in the natural gas model. Given the size of the models, it is possible to estimate savings that are less than 10% of the pre-installation billing.

τ_t is the "time-specific" error for period *t*, reflecting the unexplained difference in use between time periods,

x_{ijt} are the predictor variables reflecting the installation of energy efficiency measure *j* for household *i* in period *t*,

β_j are the slope coefficients that quantify the average influence of modeled efficiency measure *j* on monthly consumption,

p is the total number of energy efficiency measures included in the model,

z_{it} are the predictor variables reflecting nonprogram related effect *k* (such as weather impacts) for household *i* in period *t*,

γ_k represents the slope coefficients that quantify the average influence of modeled nonprogram related effect *k* on monthly consumption,

k is the total number of nonprogram related effects included in the model, and

ε_{it} is the error term that accounts for the difference between the model estimate and actual consumption for household *i* in period *t*.

The model used dummy variables, in which the *x*'s for the installed measures are one or zero to indicate the installation and the coefficients reflect the savings for the measures.

The Impact Evaluation (IE) Team reviewed the data and assessed the results to ensure that the savings estimates are statistically sound. Testing for violation of statistical assumptions was conducted for the natural gas model. The model was tested for autocorrelation[3], multicollinearity[4], outliers[5], and to assess the impacts of unequal variances across homes (heteroskedasticity).

[3] Autocorrelation of errors is most common in time-series due to the intrinsic relationship between the most recent prior period and the present measurement while unspecified variables are missing that would explain the underlying mechanisms for these changes. If the model exhibits autocorrelation, the estimators are unbiased but the variance in the model tend to be artificially low.

[4] Multicollinearity occurs when predictor variables are correlated with one another. This can happen if measures are installed as a group. If multicollinearity is present, the estimators are sometimes of the wrong sign or not statistically significant.

[5] Outliers are observations that differ significantly from the population and may have an undue influence on the results.

3.4.1 Nonprogram-Related Impacts on Energy Use

While the fixed effects model controls for the characteristics of the home that are stable over time, it is possible that the estimation of program impacts can be affected by other factors that do change over time. These types of changes can be conceptualized in two broad categories:

- Individual changes that affect specific homes, such as acquiring new household members, taking a longer vacation, or having a change in one's work schedule.
- Changes in the overall economy that affect the residential market in a global way, such as volatile gasoline prices, unemployment rates, or an increase in home heating costs.

Within-home changes may affect energy use on a house-by-house basis, but these impacts do not tend to create a bias in the final results as long as there are enough homes in the model. Within-home changes that are unrelated to the program, such as changes in plug load, are unlikely to be coincident with the measure installation and, thus, would not be expected to introduce bias into the estimated savings.

Other factors that affect energy use within residential homes, such as changes in occupancy and the addition or removal of energy-intensive equipment, create random error in the models and make it more difficult to estimate savings, resulting in estimators that are not statistically significant. If there are only a few homes in the model, these nonprogram changes could introduce bias, if, for example, there is an increase in occupancy during the post-installation period.

The current billing analyses included over 5,000 homes for the natural gas model and 3,100 homes for the electric model, which is more than sufficient to expect that nonprogram related changes would be randomly distributed over the pre- and post-periods and would not introduce bias into the analysis. In addition, the HPwES impact evaluation for PY2007/2008 included a restricted model incorporating telephone survey responses, and the results of the analysis demonstrated that the in-home changes did not create a bias in the estimated savings.

There are three common approaches to address the global factors within the statistical billing analysis:

1. Include a nonparticipant comparison group directly in the billing analysis
2. Incorporate trend lines based on consumption of the nonparticipant comparison group
3. Incorporate trend lines from third party data on critical market trends, such as unemployment rates and gas prices into the analysis.

Given that the goal of this evaluation is to estimate the RRs, the aim is to develop estimates of evaluated savings that reflect the actual reduction in energy use and do not incorporate net effects such as free riders or spillover. However, eliminating net effects is not straightforward.

The participant-only model may include net effects in the form of participant inside spillover. This type of spillover occurs when a program participant learns about efficiency through the program and then elects to install additional measures on their own at the same location. Since this net effect occurs in response to information learned through program participation, it can only occur after the program-related installation and will tend to reduce consumption during the post-installation period. To the extent that participant inside spillover occurs within the period of the billing analysis, this could result in higher savings than are actually achieved directly from the program-related measures.

While the intention of including nonparticipants to the model is to account for naturally occurring efficiency improvements or trends toward increased use, there is no effective way to separate naturally occurring efficiency from nonparticipant spillover. Nonparticipant spillover occurs when program participants spread their knowledge of efficiency practices or equipment to nonparticipants through informal conversations or other mechanisms. While nonparticipant spillover should create additional savings for the program, it would have the opposite effect in a billing model that includes nonparticipants. Consequently, incorporating nonparticipants directly into the regression model could add this net effect and reduce the evaluated savings.

In the end, a billing analysis that includes both participants and a nonparticipant comparison group will likely produce savings estimates that are somewhere in between net and gross effects and, thus, difficult to interpret with any degree of accuracy.[6] Trend lines are a less direct method of incorporating nonparticipant effects and less likely to introduce net effects. The trend line model should address the wide scale influences on energy use and mitigate some of the external effects. Incorporating trend lines reflecting nonparticipant bills will be the preferred strategy for estimating gross impacts.

The issue with introducing trend lines developed from external data such as unemployment rates or gasoline prices is that the connection between these factors and electric or natural gas

[6] This approach may be used in Phase II as an alternative approach to investigating net effects.

consumption has not been clearly demonstrated. While this approach provides an alternative to the nonparticipant trend lines, the final results are also open to interpretation.

A nonparticipant trend line was incorporated into the natural gas model. The comparison group was drawn from participants in PY 2012, using their billing data in the period prior to program participation. This step was not conducted with the electric model due to time constraints and other factors. These issues are discussed in the results section.

3.4.2 Model Selection Process

A component of the modeling process was to compare alternative models to determine the model that best fits the data and to assess the relative importance of specific variables or groups of variables. Standard statistics, such as R2 and t-values for specific parameters, were reviewed and the information-theoretic approach to model selection was employed. In conjunction, these two approaches ensured that the selection of the final model is based on objective statistical standards and the final model improves the ability to estimate the parameters of interest.

SECTION 4: RESULTS

The section presents the results and findings from Phase 1 Impact Evaluation.

4.1 GROSS ENERGY SAVINGS RESULTS

This analysis is the second billing analysis conducted for the HPwES in as many years. Table 6 compares the realization rates (RRs) for the two HPwES billing analyses. This comparison shows the results are fairly consistent, and RRs were found to be lower in the current analysis.

Table 6. Comparison of Savings for HPwES Billing Analyses

	2010-2011		2007-2008	
	Annual Electric Savings	Annual Natural Gas Savings	Annual Electric Savings	Annual Savings for All Other Fuels
RR	19%	48%	35%	65%
90% Confidence Interval	+/- 9%	+/- 1%	+/- 22%	+/- 7%
Average Evaluated Savings per Household	154 kWh	13.3 MMBtu	315 kWh	17.3 MMBtu
Percent of Pre-Installation Use Saved[1]	2%	14%	3%	16%
Number of Homes in the Model	3,185	5,009	2,536	1,462
Number of Utilities in the Model	7	7	3	3

[1] The annual consumption during the pre-installation period was averaged for all homes in the billing models. The "percent of pre-installation use saved" is the average annual evaluated savings divided by the annual average pre-installation consumptions.

To interpret the results from the electric model, the analysis was divided into measures with savings and measures with extra use. Table 7 shows the program reported and evaluated savings separately for measures with electric savings and with extra use. The electric RR is 32% for the measures with electric savings (similar to the earlier analyses), and over 120% for measures with extra electric use (primarily from heat pumps). Thus, program reported savings seem to overestimate the average savings and underestimate the extra use, which combine to create a downward effect on the RR.

Table 7. Comparison of Electric Model Results and Program Reported Savings

	Program Reported Total Savings (kWh/Year)	Program Reported Savings per Home (kWh/Year)	Evaluated Total Savings (kWh/Year)	Evaluated Savings per Home (kWh/Year)	RR	Relative Precision of Savings
Savings	3,010,860	945	951,172	299	32%	16%
Extra Energy Use[1]	(373,248)	(117)	(460,387)	(145)	123%	5%
Net Savings	2,637,612	828	490,785	154	19%	21%

[1] Extra use occurs when efficiency measures are more efficient overall but the savings are created in part by moving use from one fuel to another, generating savings of one fuel type while causing extra use in the other. For example, natural gas furnaces or oil boilers could be replaced by electric heat pumps for both heating and cooling, creating additional electric use, a reduction in natural gas use and an improvement in overall efficiency.

Extensive efforts were made to assess whether the models for program years (PYs) 2010 and 2011 were producing anomalous results, such as the following:

- Household and measure-group models were run
- Outliers and homes with very low pre-installation consumption were removed from the model
- The models were run separately for each PY
- Results were reviewed by utility
- Numerous configurations of the model variables were tried; for the natural gas model, a model that separately estimated the heating slope (therms/hdd) for each home was run (a PRISM-style approach)
- Participants with large, problematic measures (such as heat pumps) were removed
- A nonparticipant trend variable was included (for natural gas)

Although this review helped to understand and interpret the modeling results, the alternative approaches did not change the RR to any appreciable degree. When the models were run separately by PY, the RRs were quite similar. Almost all of the alternative models resulted in a RR that was the same or lower than the final model.

While the specific reasons for the lower RR cannot be ascertained directly from the billing analysis, a preliminary review of the program data identifies one difference in program implementation, in that more heat pumps were installed during PYs 2010 and 2011. The impacts of heat pumps were easily identified in the electric billing data, and the results of the modeling indicate that the additional electric use associated with these devices is substantially higher than anticipated. When the RR includes only measures with savings, it is within the confidence intervals of the two previous evaluations.

The remainder of this section describes the regression output from the models, followed by discussion of the results and supplemental analyses.

4.1.1 Regression Results

The regression output is presented in Table 8 and Table 9 below. The final models for both fuels developed estimators at the household level, *i.e.*, all of the measures were grouped into base measures[7] with savings, base measures with extra use, heating measures with savings, heating measures with extra use (for the natural gas model) and air conditioning measures (for the electric models). Further measure level distinctions did not improve the RR (or the model fit) for the electric billing model.

As natural gas use is primarily driven by heating, the modeling of heating efficiency measures in cold climates tends to be fairly straightforward. Heating use is closely related to outdoor temperature, but the characteristics of the relationship vary from one home to the next. For example, the thermostat settings, the outdoor temperature that triggers the use of the heating system and the methods of controlling the thermostat are all highly individual to specific homes. Estimating savings from base measures such as hot water conservation tends to be more complicated and the results are often more variable. For example, the inlet temperature to the water heater drops during the winter, leading to higher water heating loads, which to some extent mimics the increase in natural gas use during the heating season.

The final model uses a pooled estimate of heating use in relation to temperature variations, *i.e.*, the average heating slope for all homes was incorporated into the model. An alternative approach was tried that explicitly estimated the heating slope for each homes. This alternative model produced very similar results to the pooled model.

Table 8 below provides summary statistics from the natural gas regression model. All estimators were statistically significant at the 90% confidence interval, and the R^2 statistic was 0.82 for the natural gas model. The high R^2 statistic is due to the fixed effects model in which each home is compared to itself

[7] Base measures are not weather dependent, such as water heating conservation and appliance upgrades.

Table 8. Summary Statistics from the Natural Gas Regression Model

Measure Group	Estimator[1]	t-value[2]	Unit of Estimator	Number of Homes in Regression Model	Percent of Program Reported Savings in Model[3]
All Heating Measure Savings[4,5]	(0.024)	59.9	Therms/Heating Degree Day	4,945	95%
Base Measure Savings[6]	(0.243)	15.8	Therms/Day	837	5%
Extra Energy Use – Water Heater Fuel Switch[7]	0.302	3.1	Therms/Day	11	12%
Extra Energy Use – Space Heating System Replacement or Fuel Switch[7]	0.038	7.9	Therms/Heating Degree Day	20	88%
Heating Degree Days	0.153	550.5	kWh/Heating Degree Day	5,009	N/A
R-Square[8]		0.82			

[1] The "estimator" is the regression coefficient and reflects the impact of the variable on the change in average daily use.

[2] The t-value of a regression coefficient measures whether the value of the coefficient is statistically different from zero. The t-statistic is the regression coefficient over the its standard error. A t-value of 1.64 or higher indicates the coefficient is statistically different from zero at the 90% confidence level.

[3] The percent of program reported savings was calculated separately for savings and for extra use. Thus, the percentages for the all heating measures and base measure savings add to 100% and the percentages for the extra energy use from water heating fuel switches and space heating system replacement or fuel switches add to 100%.

[4] "Heating measures" include insulation, air sealing, heating system repair and replacement, heat pumps, programmable thermostats and other miscellaneous heating related upgrades.

[5] Interacted with heating degree days (HDD), base 60°F.

[6] "Base Measures" include measures that are not weather dependent, such as water heating conservation and replacement, and appliances.

[7] Extra use is associated with efficiency measures that are more efficient overall but the savings are created in part by moving use from one fuel to another, generating savings of one fuel type while causing extra use in the other. For example, natural gas furnaces or oil boilers could be replaced by electric heat pumps for both heating and cooling, creating additional electric use, a reduction in natural gas use and an improvement in overall efficiency.

[8] The R-squared (R^2) measures the proportion of variability in a regression data set that can be explained by the model. An R^2 of 1.0 indicates that the regression perfectly fits the data. Generally, an R^2 of 0.70 or higher reflects a strong relationship between the regression variables, from 0.30 to 0.70 reflects a moderate relationship and less than 0.30 indicates a weak relationship. A fixed effects regression as used in this analysis tends to have a high R^2 as the model compares each home to itself.

The natural gas model was stable, with the household savings reasonably consistent under different configurations of the model. Diagnostics were conducted. The Goldfeld-Quandt test for heteroskedasticity was conducted, resulting in a GQ statistics of 3.40 and indicating that the data set exhibits heteroskedascity.[8] Outliers were identified using standard statistical methods.[9] Through this process sixteen homes were removed from the analysis and the model was re-run. The results remained consistent.

In general, measure-level results tend to exhibit variability as multiple measures are installed in each home and collinearity among measures can make it difficult to determine accurate savings for specific measures, especially for measures that are infrequently installed or have small savings. Consequently, the household-level results provided in Table 9 are more reliable. However, a model distinguishing among major measure groups was run and these results are provided for informational purposes in Table 9. This analysis may be useful for identifying measures that are underperforming.

Table 9. Informational Purposes Only- Natural Gas Program Reported and Evaluated Savings by Measure Group

Measure Group	Number of Homes in Model	Program Reported Saving per Home (MMBtu/Year)	90% Confidence Limits: Lower Limit - Evaluated Savings per Home (MMBtu/Year)	90% Confidence Limits: Upper Limit - Evaluated Savings per Home (MMBtu/Year)
Base Measures[1]	837	9	5	7
Envelope (*e.g.*, insulation, air sealing, windows and doors)	3,489	23	10	11
Programmable Thermostats	1,642	7	2	4
Heat Pump	11	87	58	66
Heating System Repair/Replacement	2,145	20	6	7
Extra Energy Use – Heating System Replacement or Fuel Switch [2]	20	(75)	(13)	(20)
Extra Energy Use – Water Heater Fuel Switch[2]	11	(19)	(9)	(21)

[1] "Base Measures" include water heating conservation and replacement, and appliances.

[2] Extra energy use (negative savings) occur when efficiency measures are more efficient overall but the savings are created in part by moving use from one fuel to another, generating savings of one fuel type while causing extra use in the other. For example, natural gas furnaces or oil boilers could be replaced by electric heat pumps for both heating and cooling, creating additional electric use, a reduction in natural gas use and an improvement in overall efficiency.

---

[8] See Goldfeld in the Reference section.

[9] Outliers were identified through calculating the pooled DFFITS by household and setting a size-adjusted cut off. See Belsley in the Reference section, chapter 2.

Estimating electric savings can be more complicated as most homes do not have electric space heat and the factors driving electric use are more varied and difficult to identify. The electric model was less stable and it was not possible to estimate savings at the major measure level. However, alternative models did not improve the RR, as explained further in the Discussion section. All estimators were statistically significant at the 90% confidence interval with the exception of base extra use for heat pumps, and the R^2 statistic was 0.73 for the electric model. As with the natural gas model, the high R^2 statistic is due to the fixed effects model in which each home is compared to itself.

Diagnostics were also conducted for the electric model. The pooled Durbin Watson statistic was 1.09, indicating the presence of autocorrelation. The data set is highly likely to exhibit autocorrelation due to the consistency between billing use from one month to the next in each home. Autocorrelation does not introduce bias, but may result in artificially low variances.[10] The Goldfeld-Quandt statistic is 5.3, indicating the presence of heteroskedasticity. This statistic is approximately the same for numerous configurations of the measure variables.[11] Heteroskedasticity (unequal variances) does not introduce bias but often results in high variances.

The impacts of outliers were also estimated. About nine homes of the 3,185 were found to be outliers. When these homes were removed from the model, the overall evaluated savings went down by about 10%. As these higher saving homes are still part of the population, the full model was used to estimate the final evaluated savings. Table 10 below presents summary statistics from the electric regression model.

[10] See Sayrs in the Reference section.

[11] This result could be due to heteroskedasticity or possibly misspecification of the model. As all of the other model configurations have a similar result, model misspecification could not be due to the definitions of the measure variables.

Table 10. Summary Statistics from the Electric Regression Model

Measure Group		Estimator[1]	t-value[2]	Unit of Estimator	Number of Homes in Regression Model	Percent of Savings in Model[3]
Combined Heating Measure Savings[4,5]		(0.031)	5.14	kWh/Heating Degree Day	2,320	50%
Base Measure Savings[6]		(0.343)	1.94	kWh/Day	3,185	36%
Air Conditioning Measure Savings[7]		(0.829)	5.52	kWh/Cooling Degree Day	937	9%
Heat Pump Measure Savings[8]	Base Savings	(14.014)	14.15	kWh/Day	43	5%
	Heating Savings[5]	0.911	20.61	kWh/Heating Degree Day		
	A/C Savings[7]	(1.913)	3.65	kWh/Cooling Degree Day		
Extra Energy Use – Heat Pump[8,9]	Base Use	(0.330)	0.36	kWh/Day	58	95%
	Heating Use[5]	1.436	33.76	kWh/Heating Degree Day		
	A/C Use[7]	1.582	2.77	kWh/Cooling Degree Day		
Extra Energy Use – Water Heater Fuel Switch[8]		5.680	3.93	kWh/Day	7	5%
Heating Degree Days		0.368	12.02	kWh/Heating Degree Day	3,185	N/A
Cooling Degree Days		2.378	18.14	kWh/Cooling Degree Day	3,185	N/A
R-Squared[10]			0.73			

[1] The “estimator” is the regression coefficient and reflects the impact of the variable on the change in average daily use.

[2] The t-value of a regression coefficient measures whether the value of the coefficient is statistically different from zero. The t-statistic is the regression coefficient over its standard error. A t-value of 1.64 or higher indicates the coefficient is statistically different from zero at the 90% confidence level.

[3] The percent of program reported savings was calculated separately for savings and for extra use. Thus, the percentages for the all heating measures and base measure savings add to 100% and the percentages for the extra energy use from water heating fuel switches add to 100%.

[4] "Combined Heating Measures" include insulation, air sealing, heating system repair and replacement, programmable thermostats and other miscellaneous heating related upgrades. Heat pumps were included as a separate measure.

[5] Interacted with heating degree days (HDD), base 60°F.

[6] "Base Savings" reflect the reduction in base (non-weather-dependent) use across all homes in the model. Savings from lighting, refrigerator, water heating and appliances are likely to be driving this reduction in use.

[7] Interacted with cooling degree days (CDD), base 75°F.

[8] The heat pump measures were identified as having savings or extra use based on the Program tracking data, *i.e.,* homes receiving heat pump with electric back up were identified in the Program tracking data as having negative electric savings and are included in the “Extra Energy Use – Heat Pump” category. The regression estimators indicate that these heat pumps with electric back up resulted in increased electric use for heating and cooling, and electric base load savings.

[9] Extra use (negative savings) occur when efficiency measures are more efficient overall but the savings are created in part by moving use from one fuel to another, generating savings of one fuel type while causing extra use in the other.

[10] The R-squared (R^2) measures the proportion of variability in a regression data set that can be explained by the model. An R^2 of 1.0 indicates that the regression perfectly fits the data. Generally, an R^2 of 0.70 or higher reflects a strong relationship between the regression variables, from 0.30 to 0.70 reflects a moderate relationship and less than 0.30 indicates a weak relationship. A fixed effects regression as used in this analysis tends to have a high R^2 as the model compares each home to itself.

4.2 DISCUSSION

From a broader perspective, the results from all two billing analyses suggest that the actual, achieved savings are substantially lower than estimated by the Program. A number of supplemental analyses were conducted in response to the low RRs arising from the current billing analysis and questions posed by NYSERDA evaluation and program staff. The researchable questions are provided below.

1. How do savings from heating and cooling measures compare to the pre-installation consumption levels?
2. Why was the RR for heating system replacements radically different between the PY2007-2008 evaluation and the current PY2010-2011 analysis? As heat pumps are no longer a program measure, would the RR be different if heat pumps were removed from the analysis?
3. Were the Program reported savings closer to the evaluated savings for homes with high savings or homes with low savings?
4. Were the RRs higher or lower for homeowners who received a loan through NYSERDA?
5. Are some contractors or groups of contractors better at estimating savings than others?
6. Are Program QA efforts improving the estimates of program reported savings?
7. Did the change in program requirements used to identify cost effective opportunities have an effect on the RR?

Further analysis was conducted to address the top six issues listed above, as discussed in the following sections. Additional analysis was not conducted for the last item since the change in program procedures occurred in April of 2011, which did not leave a sufficient period for the comparison.

4.2.1 Comparison to Pre-Installation Energy Consumption for Heating and Cooling Measures

This component of the analysis was conducted to assess how the program reported savings compare to the annual energy consumption as a check on the validity of the magnitude of the program reported savings. This process involved three steps:

1. Identify the homes with heating or air conditioning consumption based on the relationship between the energy use and heating or cooling degree days.

2. Estimate the annual consumption used for heating, air conditioning and overall for each home in the model.

3. Compare the program reported average savings for heating (or air conditioning) measures to the average annual heating (or air conditioning) consumption.

If the program reported savings are a high percentage of the consumption, it suggests that savings are being overstated. The next two tables compare the program reported and evaluated electric savings for heating and air conditioning measures.

One rather unusual feature of this Program is the high percentage of the electric savings that comes from heating related measures. Figure 2 and Table 3 show the following:

- Almost half of the total program reported savings are from heating related measures, with 31% associated with envelope and heating system upgrades and 17% relating to space heating fuel switches and heat pumps.
- Most of the participating homes in the model do not rely on electricity as the primary source of heat and only 10% of homes with heating related measures used more than 4,800 kWh per year for heating, which is the equivalent of 16 MMBtu. In contrast, the average resident of New York State uses 70 MMBtu of natural gas per year to heat their homes.

Even accounting for the difference in the efficiency of electric baseboard and natural gas furnaces, it is clear that there are very few homes with substantial electric space heating consumption.

In the absence of electric space heat, the majority of these envelope and heating system electric savings are most likely due to reductions in the use of the electric auxiliary equipment, such as blower fans on furnaces or pump motors on boilers. However, average savings of 542 kWh per year seems to be a high estimate of these savings, as furnace blowers typically use in the range of 700 to 1,200 kWh per year and pumps on boilers use much less.

Table 11 compares program reported savings and heating consumption patterns for homes with electric heating measures. Of all homes in the model with envelope and heating system upgrades, there was no evidence of heating related electric use in the billing history for more than half of the homes (56%), and these homes accounted for almost half of the program reported savings. In addition, over a quarter of the program savings are associated with homes where the program reported savings were greater than 50% of the total annual heating consumption. It seems unlikely that many homes would be expected to save over 50% of their total annual electric heating use from these measures.

Table 11. Comparison of Program Reported Savings and Heating Consumption Patterns for Homes with Electric Heating Measures

	# of Homes	Total Savings (kwh)	Average Savings per Home (kWh)[1]	% of Homes	% of Total Savings
All Homes with Electric Heating Measures[2]	2,320	1,258,142	542	100%	100%
Homes without Electric Heating Use Evident in Bills[3]	1,299	595,022	458	56%	47%
Homes with Electric Heating Use Evident in Bills	1,021	663,120	649	44%	53%
Savings Less than 50% of Heating Use	*699*	*313,062*	*448*	*30%*	*25%*
Savings Greater than 50% of Heating Use	*322*	*350,058*	*1,087*	*14%*	*28%*

[1] The average savings per home were calculated by dividing the total savings by the number of homes.

[2] Only envelope and heating system upgrades were included in this table. Space heating fuel switches and heat pumps were excluded.

[3] A house-specific regression analysis was conducted for each home using the pre-installation billing records and heating and cooling degree days. Homes that showed an increase in use as the weather became colder and the total annual heating use was greater than 100 kwh were assumed to have electric heating-related use. The homes that did not show any increase in electric use as the weather became colder are identified as “homes without electric heating use evident in bills.”

Table 12 below compares savings and consumption for cooling measures. Billing records for almost a half (436) of all homes that had cooling measures installed through the Program (937) indicate that these participants had little or no air conditioning use during the pre-installation period. These participants account for about half of the program reported savings from air conditioning measures. In addition, the total savings from cooling measures were found to be greater than 50% of the total air conditioning consumption in about 5% of all homes with cooling measures accounting for 11% of the savings from air conditioning measures.

Table 12. Comparison of Savings and Consumption for Cooling Measures

	# of Homes	Total Savings (kWh)	Average Savings per Home (kWh)[1]	% of Homes	% of Total Saving s
All Homes with Air Conditioning Measures[2]	937	285,610	305	100%	100%
Homes without Air Conditioning Use Evident in Bills[3]	436	146,656	336	47%	51%
Homes with Air Conditioning Use Evident in Bills[3]	501	138,954	277	53%	49%
Savings Less than 50% of Air Conditioning Use	*456*	*108,858*	*239*	*49%*	*38%*
Savings Greater than 50% of Cooling Use	*45*	*30,096*	*669*	*5%*	*11%*

[1] The average savings per home were calculated by dividing the total savings by the number of homes.

[2]A house-specific regression analysis was conducted for each home using the pre-installation billing records and heating and cooling degree days. Homes that showed an increase in electric use in the hot weather and the total annual electric use associated with the hot weather was greater than 50 kwh were assumed to have air conditioning use. The homes that did not show any increase in electric use as the weather became hotter are identified as “homes without air conditioning use evident in bills.”

[3] Pre-installation billing records were analyzed to identify a weather dependent increase in use during the summer months.

The regression results may not present a complete picture of air conditioner use and savings because customers of Consolidated Edison (ConEd) near New York City are not included in the model and this area of the state has a higher cooling load than other regions. As the billing data provided by Con Ed covered only a portion of the analysis period, only one Con Ed home made it through the data cleaning process.

To assess the possible impacts of this omission, a supplementary analysis was conducted to examine the distribution of program reported air conditioning measures by utility. The distribution of homes with program reported air conditioning savings by utility are presented in Table 13. To provide a meaningful comparison, the analysis was restricted to the time period included in the models, *i.e.*, PY2010-2011.

Table 13. Distribution of Homes with Program Reported A/C Savings by Utility

	In Population			In Billing Model		
	Number of Homes	% of Homes	% of Savings	Number of Homes	% of Homes	% of Savings
RG&E	849	57%	49%	597	64%	57%
National Grid	340	23%	17%	165	18%	13%
NYSEG	152	10%	8%	98	10%	9%
Orange & Rockland	88	6%	16%	59	6%	18%
Con Ed	36	2%	8%	1	<1%	<1%
Cental Hudson G&E	24	2%	3%	17	2%	3%
Municipal[1]	1	<1%	<1%	0	0%	0%
Total	**1,490**			**937**		

[1] The location of the municipal utility was not specified in the program tracking data.

During the analysis period, a total of 1,490 program participants installed air conditioning measures, accounting for program claimed savings of 437,146 kWh. Customers of ConEd account for approximately 2% of the homes and 8% of the air conditioning kWh savings. Considering that air conditioning measures represent only 5% of total reported savings and that air conditioning savings from ConEd customers represent only 0.4% of the total portfolio, the magnitude of any potential bias would be exceedingly small.

In aggregate, this analysis leads to the conclusion that electric savings are being systematically overestimated for both heating and cooling measures. The most likely reason for the overstatement is the methods used to estimate savings, as other potential reasons for the low RRs, such as low installation rates or the removal of measures by the participant, are unlikely to apply to heating and air conditioning measures.

4.2.2 Heating System Replacements

The measure-level analysis conducted for the PY07/08 impact evaluation found that the RR for natural gas heating system replacements was quite high, over 100%. In contrast, the results from this evaluation suggest that the RR for these measures has substantially dropped to less than 50%. The Impact Evaluation Team conducted some additional analysis to assess potential reasons for this difference.

In general, the measure-level analysis is less reliable than the whole house regression due to the range of measures that may be installed in each home. When multiple measures are installed in the same home, there is overlap across the measure variables within the homes and the regression variables may not be able to fully separate the individual measures. In addition, there are interactive effects in the home; for example, when heating system replacements and insulation are installed, the combined savings is lower than estimated by adding the savings for the two individual measures. The overlapping measures in each home and the interactive effects make it more difficult to estimate savings for specific measures.

In both evaluations, heating system repairs and replacements were grouped as a single variable. However, a review of the PY 2010-2011 participants suggests that heating system repairs are unlikely to affect the results since heating system replacements were far more common than repairs, as shown in Table 14 below. This analysis indicates that the total magnitude of heating system repair savings is too small to produce a noticeable change in the overall RR for this measure group, even if the measure-specific RR differ widely.

Table 14. Incidence of Heating System Replacements and Repairs

Measure Category	Number of Homes in the PY 2010-2011 Model	Program Reported Savings	Percent of Homes in the PY2010-2011 Model	Percent of Program Reported Savings
Repair	33	439	2%	1%
Replacement	2,112	42,236	98%	99%
Total with Heating System Replacement or Repair	2,145	42,675		

To assess the potential impacts of overlapping measures and interactive effects, the incidence of heating system replacements in the regression models from the two evaluations was compared. This analysis showed that about 22% of the homes in the PY 2007-2008 evaluation had heating system replacements or repairs, as compared to 41% in the PY 2010-2011 evaluation. Thus, the incidence of heating system replacements and repairs almost doubled between the analysis periods of the two evaluations.

The next step was to compare the combination of measures found in the model. This comparison is shown in Table 15. In the earlier evaluation, about 8% of the homes in the model had only a heating system repair and replacement as compared to 20% in the current model. In addition, 70% of the homes with heating system measures in the earlier evaluation also installed envelope measures, as compared to 35% in the current analysis. This analysis shows a substantial difference in the combination of measures between the two evaluation periods.

Table 15. Comparison of Heating System Measure Combinations

Measure Combination	Number of Homes in Model with Heating System Replacements or Repairs[1]		Percent of All Homes in Model with Heating System Replacements or Repairs[1]	
	2010-2011	2007-2008	2010-2011	2007-2008
Heating System Replacement/Repair Only	429	27	20%	8%
Heating System Replacement/Repair and Thermostat	965	72	45%	22%
Heating System Replacement/Repair and Envelope	251	83	12%	26%
Heating System Replacement/Repair, Envelope and Thermostat	500	140	23%	44%
Total Homes with Heating System Replacements or Repairs[2]	2,145	322		

[1] Replacements account for the vast majority of the heating system upgrades during both periods.

[2] Heating system replacements include both furnace and boiler replacements. On average, program reported savings for boilers were much higher than for furnaces and the results of the billing model suggests that the boiler replacements are not achieving savings.

This analysis indicates that there were substantial differences between the two periods which may have affected the estimated savings for heating system replacements. The key points are summarized below.

- Heating system replacements were more common during PY2010-2011 and the model was larger; thus, the current analysis period included about seven times as many homes with heating system replacements (over 2,000 as compared to around 300).
- Heating system replacements were half as likely to be installed in conjunction with envelope measures in the current analysis period as compared to the prior analyses.
- Furnaces had lower program reported savings and a higher RR than boilers in the current analysis period; changes over time in the installation rates and program reported savings of furnaces and boilers could affect the RR for heating system replacements as a whole.

These results suggest that there may be a number of valid reasons for the difference in RRs for heating system replacements between the current PY2010-2011 and PY2007-2008 impact evaluations.

The natural gas model was run excluding the homes with heat pumps, and the RRs were very similar to the values reported in this memo. As part of the Phase 1 analysis, the Impact Evaluation Team will review the impact of including heat pumps in the electric model.

4.2.3 Realization Rates by Program Reported Savings Level

Program staff expressed interest in understanding whether the RRs were higher for homes with high savings. To answer this question, analysis was conducted by the level of the savings. For both the natural gas and electric models, homes were divided into four groups according to the magnitude of the program reported savings, as shown below. The savings levels were defined to generate four groups with roughly the same number of homes in each group. For natural gas, only the heating related measures were used to develop the size categories, as 92% of the program reported savings are from heating measures.

For both fuel types, the analysis shows that the RRs for the homes with lower program reported savings are substantially higher than for homes with high savings. In the lowest size stratum, the modeled natural gas savings are almost twice as high as the program reported savings, and the results from the electric model also indicate that program reported savings are being underestimated twofold. In contrast, the RRs for the homes with the highest savings are 23% and 9% for natural gas and electricity, respectively. Program reported savings, evaluated savings, and RRs are given by savings level in Table 16 and Table 17.

Table 16. Natural Gas Realization Rates by Savings Level

Savings Level	Program Reported Savings Range (MMBtu/Year/Home)[1]	Total Number of Homes in the Model	RR
1	Less than 10.4	1,216	192%
2	10.4 to 22.6	1,225	80%
3	22.6 to 39.8	1,250	47%
4	39.8 and up	1,254	23%

[1] This analysis includes only heating-related measures, as these measures account for 92% of the total program reported savings.

Table 17. Electric Realization Rates by Savings Level

Savings Level	Program Reported Savings Range (kWh/Year/Home)	Total Number of Homes in the Model	RR
1	Less than 400	732	233%
2	400 to 585	843	44%
3	585 to 1,000	781	20%
4	1,000 and up	728	9%

For both models, the regression estimators were highly significant, and the R-squared values were equivalent to that of the final models used to estimate the program evaluated savings.

4.2.4 Realization Rate by Loan Funding

A question was raised regarding whether the overstatement of savings was more pronounced for homeowners who were planning to obtain a loan from NYSERDA. To investigate this possibility, an additional model was run to estimate the RRs for homeowners who obtained a NYSERDA loan.

Since the heating measures typically installed in homes with natural gas constitute a large majority of the program reported energy savings and costs, this analysis was conducted for the natural gas heating model only. Category definitions for the alternative model are provided with the results in Table 18.

Table 18. Natural Gas Savings and Realization Rates by Loan Source

NYSERDA Loan	Number of Homes in the Model	Program Reported Savings per Home (MMBtu/Year)	Evaluated Savings per Home (MMBtu/Year)	RR for Heating Measures
Energy $mart Loan	376	40.5	17.4	43%
Green Jobs-Green New York Loan	357	45.3	15.7	35%
No Program Loan	4,212	24.5	11.3	46%

Homeowners with GJGNY loans have a slightly lower RR than the participants who did not receive a loan, and this difference is statistically significant at the 90% confidence level. However, given the small number of participants with GJGNY loans, this difference does not have a substantial effect on the overall program RR.

4.2.5 Realization Rate by Contractor

Another request from program staff was to identify whether the RRs vary substantially among the major contractors and whether the length of time a contractor has worked with the program may affect RRs. Two additional natural gas models were run to assess the contractor-specific RRs. The first approach was to identify the homes completed by the largest contractors and estimate RRs for the ten contractors with the most completed projects. The remaining homes were aggregated into two strata to reflect medium-sized and smaller contractors. The second analysis divided homes based on the length of time the contractor had been working within the HPwES program. These additional analyses were conducted for the natural gas model only, as most of the program savings are from natural gas and other fossil fuels.

A third analysis was conducted to assess whether the Program's quality assurance efforts in place during 2010 and 2011 may be improving RRs.

The final set of 5,009 homes included in the final natural gas model formed the starting point. However, to allow a direct comparison between contractors, a further data cleaning step was added: any participant homes with work completed by more than one contractor were excluded from the models. As the vast majority of participant homes were not in this category, consequently, this filter, which reduced the total number of homes in the models to 4,789, is not expected to introduce any bias. Contractor category definitions for each model are provided below. The "Top 10" contractor category reflects the largest ten contractors based on the number of projects completed through the program during the analysis period; the RR for each of these top 10 contractors is listed separately, though none of the contractors are identified by name. Table 19 presents the RRs by contractor.

Table 19. Realization Rates by Contractor

Contractor Size Category	Number of Contractors in Analysis	Percent of Total Program Reported Savings	RR
Top 10	1	4%	82%
Top 10	1	5%	57%
Top 10	1	8%	41%
Top 10	1	4%	38%
Top 10	1	5%	53%
Top 10	1	5%	55%
Top 10	1	5%	44%
Top 10	1	4%	43%
Top 10	1	2%	92%
Top 10	1	3%	27%
Medium (60 to 175 projects)	15	25%	40%
Small (less than 60 projects)	127	28%	39%

As can be seen in Table 19 above, there is significant variation in the RRs among the top ten largest contractors, ranging from a low of 27% to a high of 92%. In aggregate, the RR for the largest ten contractors combined is 50%, which is higher than the RRs for the smaller contractors (around 40%).

Table 20 shows the RRs according to the length of time that the contractor has participated in the program. With the exception of the contractors who participated for less than two years, the difference in average RRs is statistically significant at the 90% confidence interval.[12]

Table 20. Realization Rates by Contractors' Length of Program Participation

Contractor Category	Projects Completed During Analysis Period	Number of Contractors in Category	Total Number of Homes in Model	Percent of Total Program Reported Savings	RR
Less than 2 years	110	22	100	2%	39%
2 to 5 years	1,746	63	1549	29%	40%
5 to 10 years	2,523	52	2159	48%	44%
More than 10 years	1,228	15	981	20%	53%

4.2.6 Program QA Efforts

Program staff was also interested in learning whether their quality assurance efforts were making a difference. To address this issue, the program staff provided a list of all of the Quality Assurance (QA) inspections conducted during 2010 and 2011 and the results of the inspections. For each of the top ten largest contractors listed in Table 19, the percent of passed inspections was compared to the RR and the Pearson correlation coefficient was calculated. This analysis found a correlation of 0.64, suggesting that there is a relatively strong and positive correlation between the two. This result indicates that program QA efforts are likely to be effective in improving RRs over time. This analysis, in conjunction with the contractor analysis above and data collected through other mechanisms, may provide additional insight for quality assurance activities at the program level.

4.2.7 Summary of Results

While the RRs varied considerably over the three billing analyses, the savings as a percent of pre-installation use is reasonably consistent for natural gas. Table 21 shows the annual consumption during the pre-installation period and the evaluated savings as a percent of the annual consumption, as calculated from the billing models. Comparing the previous evaluation results

[12] The realization rates are based on the heating measures, which account for the vast majority of the program reported savings. The regression estimators from both of the supplemental billing models were highly significant, and the R-squared values were 0.82 in each case, equivalent to that of the final natural gas model.

for PY2007-2008 to the current evaluation for PY2010-2011, the natural gas savings accounted for 16% and 14% of annual consumption, respectively.

Table 21. Comparison of Annual Consumption and Evaluated Savings

	Annual Consumption Prior to Program Participation		Evaluated Program Savings			
			Annual Savings per Home		% of Pre-Installation Annual Consumption	
	2010-2011	2007-2008	2010-2011	2007-2008	2010-2011	2007-2008
Annual Electric Consumption and Savings (kWh)	9,310	8,700	154	315	2%	4%
Annual Natural Gas Consumption and Savings (therms)	960	1,055	133	173	14%	16%

An impact evaluation of the Home Performance with Energy Star (HPwES) program in Vermont[13] was recently completed and provided results for electric and fossil fuel (propane, oil, and kerosene) savings.[14] The fossil fuel results are almost identical to the findings of the current PY 2010 to 2011 billing analysis, with a RR of 51% and program savings of 14% of pre-installation consumption. The electric analysis for the Vermont HPwES program showed stronger savings, with a RR of 86% and household savings of 384 kWh per year.

[13] *Efficiency Vermont's Home Performance with ENERGY STAR® Program Impact Evaluation, Final Report*. Prepared for the Vermont Department of Public Service. May, 2013. West Hill Energy and Computing, Inc. with GDS Associates, Inc.

[14] The fossil fuel savings were unregulated fuels, such as propane, oil and kerosene. As part of the impact evaluation, the results were supported by an alternative analysis for homes with natural gas, although the sample size was small (76).

SECTION 5: CONCLUSIONS

The results from the three billing analyses conducted over different time periods and incorporating billing records from various utilities all suggest that program savings are being substantially overestimated. Supplemental billing analyses provide some insights that assist with identifying potential issues with program reported savings, as discussed below:

- Comparing program reported savings from heating or cooling measures to the annual heating or cooling consumption during the pre-installation period suggests that program reported savings for these measures are overstated for many homes.
- The impacts of heat pumps were easily identified in the electric billing data, and the results of the modeling indicate the additional electric use associated with these devices is substantially higher than anticipated.
- Realization rates (RRs) vary substantially by contractor, and contractors who have been participating in the program longer have a higher RR than contractors who are newer to the program.
- On average, homes with lower program reported savings tend to have a higher RR than homes with higher program reported savings.
- While the RR for homeowners who received a GJGNY loan was somewhat lower than the RR for homes without a loan (35% as compared to 46%), the small number of participants with a GJGNY loan indicate that this factor is unlikely to be a major contributor to the overall low program RR.
- There is a relatively strong and positive correlation between the percent of inspections passed and the RR for the largest ten contractors, suggesting that program QA efforts are having a positive impact on improving the accuracy of program reported savings.

While the RRs vary among the recent three billing analysis, the evaluated savings as a percent of the pre-installation consumption is fairly consistent for natural gas (14% in the current PY 2010-2011 evaluation and 16% in the previous PY 2007-2008 evaluation).

SECTION 6: REFERENCES

Belsley, D.A., Kuh, E., and Welsch, R.E. *Regression Diagnostics,* New York: John Wiley & Sons, Inc., 1980.

Burnham, Kenneth and David Anderson. 2002. *Model selection and multimodel inference.* New York: Springer-Verlag.

The California Evaluation Framework. TecMarket Works, et. al. Project Number: K2033910. Prepared for the California Public Utilities Commission and the Project Advisory Group. September, 2004

California Energy Efficiency Evaluation Protocols: Technical, Methodological and Reporting Requirements for Evaluation Professionals. Prepared for the California Public Utilities Commission. April, 2006.

Dielman, Terry E. 1989. *Pooled cross-sectional and time series data analysis.* New York: Marcel Dekker.

Efficiency Vermont's Home Performance with ENERGY STAR® Program Impact Evaluation, Final Report. Prepared for the Vermont Department of Public Service. May, 2013. West Hill Energy and Computing, Inc. with GDS Associates, Inc. (EVT HPwES)

Goldfeld, S.M. and R.E. Quandt. 1965. *Some tests for homoscedasticity. Journal of the American Statistical Association* 60 (2): 539-547.

MassSAVE Final Summary QA/QC and Impact Study Report. Prepared for Bay State Gas, Berkshire Gas, Cape Light Compact, KeySpan Energy Delivery, National Grid, NSTAR Electric and Gas Corporation, Unitil/Fitchburg Gas and Electric, New England Gas Company and Northeast Utilities/Western Massachusetts Electric. Prepared by RLW Analytics, Middletown, CT. April 8, 2008. http://www.env.state.ma.us/dpu/docs/electric/08-46/82908nsteera6s9.pdf (MASS)

NYSERDA 2007-2008 EMPOWER NEW YORK^{SM} Program Impact Evaluation Report, prepared for the New York State Energy Research and Development Authority. Prepared by Megdal & Associates, LLC. February 2012.

"Recommendations for Targeted Home Performance with Energy Star Energy Impacts" (Memorandum) from PA Consulting Group to Oscar Bloch, Public Service Commission of Wisconsin. June 28, 2009. http://www.focusonenergy.com/files/Document_Management_System/Evaluation/Recommendations_for_Targeted_Home_Performance_with_ENERGY_STAR_Energy_Impacts_-_Revised_Final_v2.pdf (WIS)

Report on Impact Assessments of Two NYSERDA Natural Gas Efficiency Programs: Con Edison Natural Gas Efficiency Program and National Grid Low Income Natural Gas Efficiency Program, prepared for the New York State Energy Research and Development Authority by ERS, North Andover, Massachusetts. June, 2010 (NYSERDA)

Sayrs, Lois W. 1989. *Pooled time series analysis*. Newbury Park, California: Sage Publication