Home Performance with ENERGY STAR® Program Impact Evaluation Report

Final Report

Volume 3: Phase 2 Investigation into Program Savings

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 3 presents the Phase 2 Investigation into Program Savings. This component of the evaluation was designed to evaluate the potential reasons for the overestimate of savings found in previous evaluations of HPwES, including the Phase 1 billing analysis. The first component of the evaluation was conducted using the HPwES quality assurance records and contractors' modeling files for selected homes. The second component was a review of the contractor survey done by the Process Evaluation team.

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 to understand program procedures and policies. In addition, the ERS team appreciated 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

LIST OF TABLES

LIST OF FIGURES

SECTION 1: INTRODUCTION

Volume 3 covers the Phase 2 Investigation into Program Savings. The Impact Evaluation Team conducted four separate billing analyses to verify the natural gas savings from the HPwES Program covering program years 2007 through 2011. These analyses indicated that program reported savings were substantially overstated.

This component was designed to investigate the potential reasons for the over-estimation of savings. A list of possible causes was constructed and the evidence was reviewed to support or eliminate the sources of the discrepancy. This analysis covered the following five areas:

1. Model inputs
2. Calibration to pre-installation consumption
3. Software algorithms
4. Errors in data transfer
5. Quality of the installations

This evaluation used the data collected by the contractors during the audit and modeling process rather than conducting new site visits. This analysis was intended to expand on existing analyses, both by looking at more projects and by conducting a more in-depth look at the project files. In some cases, there was insufficient evidence to come to a firm conclusion.

As a second part of the program savings analysis the Impact Evaluation Team reviewed the results of contractor surveys completed by the Process Evaluation Team. This review was the result of the following process:

1. Contractor interviews were planned by both the Process Evaluation and the Impact Evaluation teams in the current evaluation cycle.
2. The Process Evaluation Team fielded their survey in September and October, 2014.
3. NYSERDA requested that the Impact Evaluation Team review the results from the Process Evaluation survey for two purposes:
   a. To gather information that may be relevant to understanding why the program realization rates (RRs) are low
   b. To identify other areas of research that could be useful for understanding why the RRs are low

This review was used to further inform the results of the contractor collected data.

1.1 PROGRAM DESCRIPTION

HPwES encourages home and building owners and tenants of existing one- to four-family homes and small low rise buildings to implement comprehensive energy efficiency-related improvements. Improvements and technologies are installed by participating contractors accredited by the Building Performance Institute (BPI). Eligible measures include building shell measures, such as air sealing and insulation; efficient appliances, such as ENERGY STAR refrigerators; heating measures, such as boilers and furnaces; cooling measures, such as ENERGY STAR room or central air conditioners; water conservation measures; domestic hot water improvements; efficient lighting; and certain renewable energy technologies.

The comprehensive home energy assessment has been the foundation of HPwES from its inception. Free energy assessments were offered through the Green Jobs-Green New York (GJGNY) starting in November 2010, and participating in the HPwES program involved the following sequential steps:

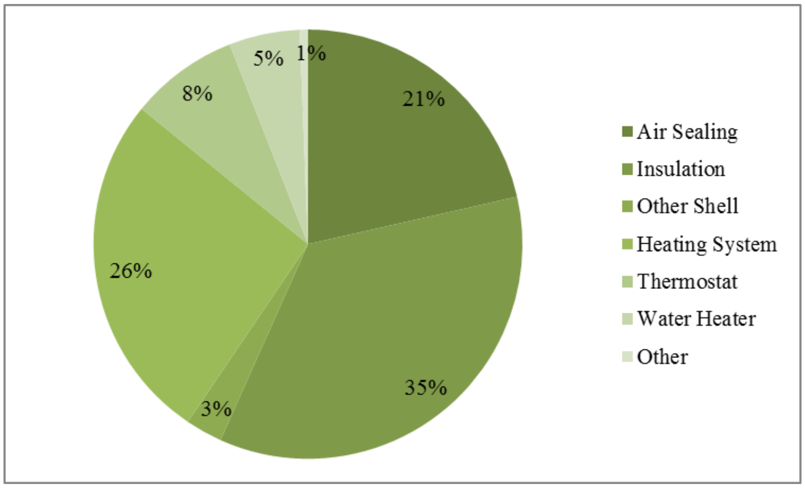
- An interested household requests a home energy audit through the GJGNY
- A participating HPwES contractor carries out the home energy audit
- Based on the audit, the participating contractor creates a building model in TREAT or HomeCheck and uses the model to quantify the energy impacts of a package of recommended efficiency measures; this model is uploaded to the program portal as “audit”
- The participating contractor provides an audit report and recommendations to the participant household
- The participant household decides which efficiency measures, if any, they are interested in having installed and contracts with either the contractor who completed the audit or another HPwES participating contractor to complete the work
- The selected contractor uploads an updated model of the home that includes the selected package of efficiency measures as “contract”
- The selected contractor carries out the work and updates the home model a final time to reflect any changes from the contracted scope of work as well as any post-installation measurements of performance (e.g., CFM50 from a blower door test) as “complete”

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

A wide range of energy efficiency measures are installed through the Program, from screw-in 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.

Program tracking by measure is highly detailed and both savings and extra use (which occasionally results when measures switch energy use between two fuel types, such as replacing electric baseboard heat with a natural gas boiler) are recorded for each fuel type. Figure 1 shows the natural gas savings by measure category.

Figure 1: Program Savings by Measure Category 2010-2011



Changes have been made in HPwES protocols over time to try to improve the modeling. In 2011, several validation rules were added to the database to check the models being uploaded. In particular, two of the rules could have an impact on the contractors' inputs:

- Reject models with wall cavities of less than R-3
- Reject models with ceiling cavities of less than R-5

In addition, the Green Jobs-Green New York audit became the primary mechanism for entering the HPwES Program starting at the end of 2010. As part of this audit process, participants were required to provide the billing records for heating fuels.

SECTION 2: METHODS

This section describes the data sources and analysis methods used to investigate each of the hypotheses for low program savings.

2.1 OVERVIEW

The objective of this component of the evaluation was to provide insight into the reasons for the overstatement of savings seen in the results of previous HPwES billing analyses. The five hypotheses for the overstatement of savings, as outlined in the work plan, are as follows:

1. **Model inputs**

   The model inputs are inaccurate, resulting in overestimated savings; for example, the pre-installation conditions were modeled as being less efficient than was actually the case.

2. **Software algorithms**

   TREAT model algorithms are not accurately estimating the impacts of some measures, such as the interactive effects of installing shell and heating system upgrades.

3. **Calibration to pre-installation consumption**

   Baseline household models are not being reconciled to actual utility bills; modeling results that are calibrated to bills are more likely to be accurate on average.

4. **Errors in data transfers**

   Problems in importing modeled savings from TREAT to the program tracking database lead to discrepancies in the program reported savings.

5. **Quality of the installations**

   Measures are not being installed correctly, resulting in lower savings than anticipated from modeling.

Each of these potential sources of error was investigated to the extent possible with the available data. The data sources and a description of the analytical details are included below.

The review of the contractor survey the primary research design included an investigation into installation practices and other aspects of contractor engagement in HPwES. Five major areas for investigation were identified:

1. Hiring and training practices
2. Internal quality assurance/quality control (QA/QC)
3. Modeling and use of participant consumption data
4. Diagnostic testing
5. Installation practices

In addition, two other issues of lessor importance were identified: the use of subcontractors, which is encouraged by the Program, and consistency of installation practices between HPwES and non-HPwES projects.

2.2 DATA SOURCES

Three data sources were used in this evaluation:

- QA records of projects inspected between 2010 and 2014
- Model output files from projects analyzed in TREAT
- Natural gas consumption data from the Phase 1 analysis of 2010 and 2011 projects.

For this component of the Phase 2 evaluation, 630 projects were randomly selected from projects that had a QA inspection between 2010 and 2014. Both TREAT[1] and HomeCheck are used to model the homes, but the majority of projects are analyzed using TREAT, so the investigation focused on projects modeled with TREAT. As 12% of the selected files were modeled with HomeCheck, the remaining 563 had TREAT output files and were used in the analysis.

The primary source of data used in the analysis is the TREAT output files for the selected projects. For each home, three TREAT output files in XML format were available summarizing

1) the audit results
2) the contract with the participant
3) the completed version reflecting the actual installation

---

[1] Performance Systems Development Consulting. TREAT. Computer software. Vers. 3.4. N.p., Aug 6, 2015. Web. <http://psdconsulting.com/software/treat>. Note: TREAT has been licensed by Performance Systems Development Consulting since 2002. Version 3.4 is the latest release from Oct 2014. Many contractors continue using earlier versions.

In addition, the Evaluation Team had access the current version of TREAT and used it to get a better understanding of the data from the XML files and how the software was used.

Data cleaning of the modeling files was straightforward, as the XML output files were already in a standardized format. To allow statistical analysis of the entire set of files, Analytical Evaluation Consultants developed a program to generate a combined database for each set: audit, contract, and completed projects. The majority of the analysis was done using the XML output files for the completed projects as that is the basis of the claimed savings.

The other data used in this evaluation were pre-installation natural gas billing records collected and analyzed as a part of Phase 1. The billing data were only available for 125 of the projects from PY10 and PY11. For these projects, the modeled consumption from TREAT was compared to the actual consumption from the billing data analysis. For the remaining projects, only the TREAT output data were available.

For the review of the contractor survey the only data source was the Process Evaluation contractor survey memo and the survey questions used.

2.3 SAMPLING

For this evaluation, 100 projects from each year between 2010 and 2012 and 150 projects from 2013 and 2014 were randomly selected from the QA inspections completed in those years. An additional 30 projects from 2010 and 2011 that had undergone administrative review were selected in the hopes they would have TREAT modeling files available for review.[2] Projects are selected to have a QA inspection for one of several reasons, as described below:

- When a new contractor joins the program, the first three projects receive a QA inspection.
- At least one project per contractor is inspected each year, with a goal of 15% of their projects inspected each year. More inspections may be done based on past performance.
- The homeowner may request an inspection, usually because of a complaint.

Since the QA inspections are focused largely on projects conducted by new contractors, projects with participant complaints and contractors with historical issues, it is likely that projects with installation problems are overrepresented in the sample. Consequently, the sample may not be

[2] The Impact Evaluation Team later learned that TREAT modeling files were not available.

representative of the program as a whole. However, this approach may be more useful for identifying and understanding problems with the program reported savings estimates.

2.4 ANALYSIS

2.4.1 XML File Analysis

Each of the five hypotheses was analyzed separately, as described below.

1. Model Inputs

The model inputs were analyzed with a review of the TREAT XML files. The average inputs were reviewed to assess if the inputs were within a reasonable range. The natural gas billing data was used to compare the high and low performing projects and any differences that may have led to low realization rates.

2. Software Algorithms

It was not possible to conduct a thorough investigation into the TREAT software as the Impact Evaluation Team did not have access to the underlying algorithms. In addition, NYSERDA's HPwES Program Staff recently commissioned a study to investigate how TREAT compared to other similar products.[3] The Impact Evaluation Team focused on interactive effects between measures and whether the modelling is less reliable for older homes. The interactive effects were reviewed by comparing the savings in TREAT for individual measures and the combined project savings to determine the magnitude and direction of the interactive effects. In addition, key metrics were compared for older and newer homes.

3. Calibration to Pre-Installation Consumption

The natural gas billing data were used to assess if project savings were compared to pre-installation consumption. As one of the model inputs was the number of utility bills entered into TREAT, projects with bills were compared to projects without bills.

[3] NYSERDA Home Performance with Energy Star Realization Rate Attribution Study. Prepared by Performance Systems Development. January 2015.

4. Errors in Data Transfers

This analysis was designed to assess whether errors in transferring the savings from the contractors' models to the program database could be contributing to the low realization rates. To do this, program savings from the database were compared to the TREAT savings for each measure.

5. Quality of the Installation

Site visits were outside the scope of the evaluation. The Impact Evaluation Team investigated whether the narrative descriptions from QA site visits provided sufficient detail to assess issues with the quality of the installations. However, the descriptions contained limited notes about the installation issues. The Impact Evaluation Team carefully reviewed the notes for each project and separated the installation issues that could affect savings into categories. There was insufficient detail in the notes to quantify the impact of each type of issue.

To provide additional insight into effects that did not fall within the five hypotheses, several other trends were evaluated using the model input averages:

- changes in inputs and modeled consumption across program years
- differences among contractors
- project savings to investment ratio (SIR)

These additional trends were analyzed by comparing averages of the TREAT XML file inputs across the relevant categories (program years, contractors, different SIR values). The natural gas billing data was also used to compare differences in performance across groups.

2.4.2 Contractor Survey Review

The contractor survey review was done by identifying the questions and results of the Process Evaluation survey that were relevant to the five areas listed above. Each of the areas was reviewed and missing questions were identified for future investigation.

SECTION 3: RESULTS

An initial review was conducted to distinguish between projects more likely to have substantially overstated savings and those that were less likely. The key metric was the savings fraction, i.e., the percent of the pre-installation consumption assumed to be saved by the installed measures. The evaluated savings indicate that the average savings fraction is in the range of 15 to 18%. However, the savings fraction based on the program reported savings is often much higher.

The savings fraction was calculated in two ways:

1. The program savings fraction (PSF), the program savings divided by the TREAT modeled consumption)
2. The billing savings fraction (BSF), the program savings divided by the consumption from the billing data)

While the PSF was calculated for all 563 projects in the analysis, the BSF could only be calculated for the 125 PY2010 and PY2011 projects with available billing data. As billing data was not available for the later projects, it was not possible to compare trends over time.

- A comparison of the modeled and actual annual consumption, BSF, size of the home and median realization rates is shown in Table 1, and illustrates the following:
- There is a strong relationship between the average BSF and the ratio of modeled to actual pre-installation energy use.
- In all quartiles, the modeled pre-install consumption is substantially higher (at least 25%) than the actual consumption estimated from bills, and in the top quartile, the modeled annual use is more than twice the actual use.
- The BSF and the median RR have an inverse relationship, indicating that projects with a high savings fraction also have overstated savings.
- The homes in the highest quartile use substantially less than natural gas on average than the homes in the other quartiles (about 750 as compared to about 1,000), suggesting that
    - o the modeled pre-installation use is particularly overstated for smaller homes and homes using a secondary fuel (22% in the highest savings fraction quartile as compared to 3-13% for other quartiles), and /or
    - o the efficiency of these homes during the pre-installation period is substantially understated in the models.

This initial analysis suggests that the pre-installation consumption modeled in the TREAT software is overestimated for most participating homes, and there is a strong correlation between the overestimation of annual consumption and the overestimation of savings. Three of the hypotheses discussed above could contribute to this outcome:

1. Incorrect model inputs (leading to an underestimation of the pre-installation efficiency levels or attribution of savings to natural gas rather than the another heating fuel)
2. TREAT software algorithms
3. No calibration of the model to actual billing records

Comparison of the "low savers" (BSF quartile 1) to the "high savers" (BSF quartile 4) provides some insight into the possible reasons that savings are overestimated. Table 1 compares the annual consumption and billing savings fraction for PY2010 and 2011. The specific analyses directly relating to the listed factors is described in more detail below.

Table 1: Annual Consumption and Billing Savings Fraction for PY2010 and 2011

Billing Savings Fraction Quartile	Billing Savings Fraction[1]	Average Size of Home (sq. ft.)	Median Realization Rate[2]	Comments	
				Ratio of Modeled to Actual Consumption[3]	Pre-Install Use
1	0.14	1,800	0.57	Modeled higher by 25%	Higher than NY average by 80 therms or 9%
2	0.27	2,208	0.40	Modeled higher by 35%	Higher than NY average by 220 therms or 24%
3	0.42	1,664	0.38	Modeled higher by 50%	Higher than NY average by 90 therms or 10%
4	0.71	1,555	0.21	Modeled higher by 115%	*Lower* than NY average by 150 therms or 17%

[1] This analysis was limited to projects that received QA site visits and is not representative of the Program as a whole.
[2] A rough estimate of the realization rate was calculated for each home by comparing the energy use before and after the upgrades. This method was intended to provide some insight into the connection between the BSF and overstatement of savings. Due to the wide variation in energy use in individual homes, it is not a reliable approach to estimating evaluated savings for the program as a whole.
[3] This column compares the modeled annual consumption from the TREAT software to the actual annual consumption from the bills.

3.1 MODEL INPUTS

This hypothesis presumes that contractors are entering incorrect information into the modeling software, possibly due to no access to utility bills, lack of attention to detail or to allow them to meet the program standards for offering incentives. The initial analysis suggests that the efficiency of the homes under the pre-installation conditions could be understated, and the investigation into the model inputs largely focused on this possibility. Investigating the model inputs consisted of three analyses of the pre-installation conditions:

1. Modeling errors generated in the TREAT software
2. Insulation levels
3. Heating system efficiencies

These areas were selected as the TREAT warnings indicate where problems were previously identified and insulation and heating system replacements account for 61% of the program reported savings (See Figure 1). As air sealing accounts for another 21% of the program reported savings, the Impact Evaluation Team also considered the feasibility of assessing the air sealing inputs. The results from each of these analyses are described in more detail below.

3.1.1 TREAT Modeling Warning Messages

As part of the modeling process, TREAT generates warnings to identify potential problems with the inputs to the building models. The contractors may either correct the inputs (if needed) to stop the warnings or ignore them if the inputs were accurate. The most common warnings are shown by category in Table 2 below with the direction of the effect on the savings estimate.

Table 2: Summary of Common TREAT Warnings

TREAT Warning	Effect on Savings Estimate		Base Building (Number of Homes)	Percent of Homes
	Over-estimation	Under-estimation		
High heating system capacity[1]		X	189	34%
Low heating system capacity	X		95	17%
Surfaces with R-Value < R-4	X		219	39%
Unusually low window area for space size[2]	X		316	56%

[1] It is unclear why a high heating system capacity would lead to an underestimation of savings. The Impact Evaluation Team tried a number of scenarios in TREAT for several homes, and the results were consistent.
[2] Underestimating the window area would lead to an overestimation of savings for insulation measures, as it artificially increases the wall area and, thus, the area that could be insulated.

These warnings do not necessarily mean there are errors, but they do indicate that the inputs or outputs are unusual. Overall, 75% of homes in the analysis had one or more errors that suggest savings could be overstated.

In general, the incidence of these warnings is not significantly different between the low and high savers with the exception of the heating system capacity. Among high savers, there are more projects with low heating system capacity. Low savers had more projects with warnings relating to high heating system capacity. A comparison of the TREAT warnings for the base home and the installed measure scenario indicates that the low R-value warning often occurs both before and after the work is done, suggesting potential missed opportunities.

3.1.2 Pre-Installation R-Values

A small difference in pre-installation R-values leads to a large difference in both the annual consumption and predicted savings. To assess whether the input R-values for the pre-installation conditions appear to be reasonable, the R-values for the low and high savers (see Figure 2) were compared. To allow inclusion of all the projects, the PSF quartiles were used for this comparison rather than the BSF.[4]

In the absence of site visits, this analysis does not definitively demonstrate that the model inputs were incorrect. However, the findings suggest that pre-install R-values may be understated, as described below.

- The differences in attic and wall median R-values between the high and low savers were substantial, at 5.8 (high savers) v. 10.0 (low savers) for attic insulation and 5.9 v. 10.0 for walls, as shown in Figure 2. The savings for an attic insulated to R-43 are 49% higher if the attic had an R-value of 5.8 prior to the installation, as compared to an attic with an R-value of 10.0.
- In TREAT, a 2x4 wood frame wall with no insulation is modeled as an R-value of 4.4, suggesting that many of the high saver homes were modeled as completely uninsulated.

[4] The BSF and PSF are strongly correlated with a Pearson correlation coefficient of 0.80. Using the PSF allows all 563 projects to be included in the analysis rather than only the 125 PY2010 and PY2011 projects with billing data.

Figure 2: R-Value by Savings for Low and High Savers

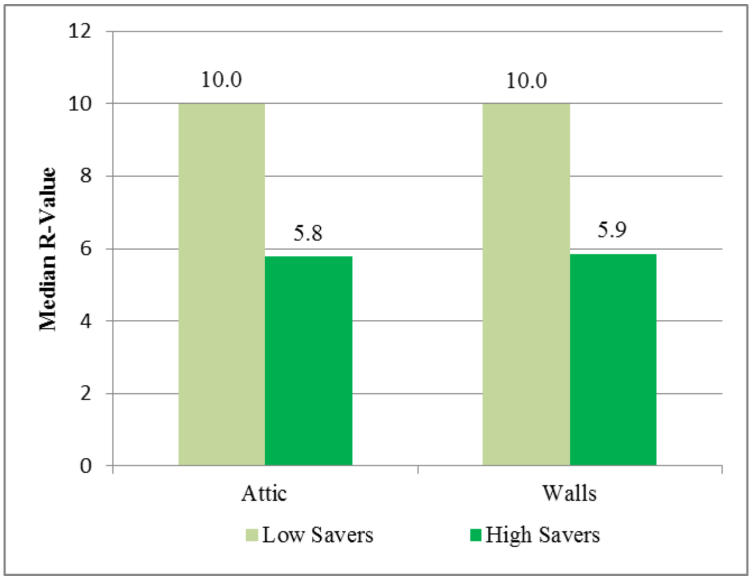


Table 3 presents the pre-installation attic R-value by age of the house. As shown in the table, while some older houses would be expected to have no insulation, the low values are not limited to the old houses. The median attic R-values for homes built before 1950 and those built between 1950 and 1970 are statistically the same at around R-7. Newer homes built in 1970 or later have a median of R-10, which seems unexpectedly low as homes were commonly built with substantial levels of insulation during this period.

Table 3: Pre-Installation Attic R-Value by Age of the House

Age of Homes	Number of Homes	Median Pre-Installation Attic R-Value	Comment
Before 1950	240	6.8[a]	Many homes were built without insulation during this period.
1950 to 1970	191	7.0[a]	Moderate levels of insulation were regularly installed.
1970 and later	132	10.2	Substantial insulation became standard practice.

[a] The difference between R-6.8 and R-7.0 is not statistically significant.

In combination, these results suggest the pre-installation R-value were understated for at least some of the homes. For wall insulation, the program guideline specific a maximum initial insulation R-value of 5 to qualify for HEMI or Assisted Home Performance Subsidy, which may encourage contractors to underestimate the pre-install R-values.

Understating the pre-installation R-values could be a major contributing cause to the difference between the billing data and TREAT consumption. The lower initial R-values will result in increased consumption in the TREAT model and higher savings when an insulation measure is installed.

3.1.3 Heating System Annual Efficiency

The efficiency of the heating system is a key input that affects the pre-installation annual consumption and the savings from both heating system and envelope measures, and the heat system savings are sensitive to small changes. For example, using a default efficiency of 80% for the existing heating equipment prior to replacement rather than the actual efficiency of 82.5% overstates the savings from a new, efficient heating system by 38%, as shown in Table 4 below. This understatement of the heating system efficiency would also result in a small overstatement of the annual consumption and savings from insulation and air sealing measures.

Table 4: Heating System Efficiency Assumption Impacts

Scenario	Pre-Installation Efficiency	New Efficient System	Percent Improvement	Overstatement of Savings
Default base case	80.0%	90.0%	13%	38%
Measured base case	82.5%	90.0%	9%	0%

For many homes, the heating system annual efficiency appears to be estimated rather than measured values. There are two patterns in the data that suggest the input values may be assumed rather than directly measured:

- Over half of the heating systems have an efficiency that is a multiple of 5% (between 60% and 90%) and 80% was the most commonly selected efficiency with 19% of all heating systems.
- Among the efficiencies that are not a multiple of 5%, the TREAT default values are common, accounting for an additional 10% of the heating systems. For example, 82% is the default for an induced draft oil or natural gas furnace from the 1990s.

Overall, about 60% of the projects have a heating system that fall into one of these two categories. Efficiencies that are TREAT defaults and multiples of 5% do not necessarily mean that they are assumed, but it seems highly unlikely that such a substantial percentage of heating systems would be either TREAT defaults or multiples of 5% if the contactors were entering the results of combustion tests.

The Impact Evaluation Team also considered whether the assumptions about heating system efficiencies could be contributing to the low RRs. The first part of this analysis was to compare the input efficiencies to the federal standards enacted in 1991, which set the floor for heating system efficiencies at 78% for furnaces and 80% for boilers. Of the projects in the sample, there were a total of 125 homes with boilers and 348 with furnaces and complete heating system data[5]. A review of the PY 2010 to 2013 projects with heating systems installed in 1993 or later shows the following:

- 26% (23) of heating systems installed between 1993 and 1999 did not meet the federal standard.
- 10% (33) of heating systems installed after 1993 did not meet the federal standard.
- The incidence of boilers that do not meet the federal standards was consistently higher than the incidence of furnaces (46% as compared to 19% for heating systems installed between 1993 and 1999).

Given that the percent of furnaces and boilers below the federal standards is substantially higher than expected, it seems likely the assumed heating system efficiencies are understated for some homes.

The final part of this analysis was to compare the input heating system efficiencies of the low savers to the high savers. This comparison is shown in Table 5 below. Among the high savers, the incidence of 80% efficient heating systems is over two times higher than for low savers (38% to 16%) and also as compared all projects (38% to 19%). These differences are statistically different at the 90% confidence level. Other comparisons did not show statistically significant differences.

Table 5: Heating System Efficiency Assumptions in High and Low Savers

Group	Number of Projects	Projects with Heating System Efficiency of 80%			
		Number of Projects	Percent of Projects	Lower Confidence Limit	Upper Confidence Limit
Low Savers	31	5	16%	5%	27%
High Savers	32	12	38%	23%	52%
All Projects	563	106	19%	16%	22%

[5] The audit data collected included an installation year and efficiency for the existing baseline heating systems in place at the time of the audit.

This analysis suggests that assumed efficiencies are more common among the high savers than the low savers and may be a contributing factor to the low RR's.

3.1.4 Air Sealing

Air sealing is a major measure and accounts for 21% of the program reported savings. The impacts of air sealing are based on the initial and final CFM50 values as measured during the blower door test conducted by the contractor. Higher reductions in CFM50 values indicate more potential savings. The measured reductions are not, however, a direct measurement of savings.

Investigation of the CFM50 values showed a difference between the high and low savers, with the high savers having a higher initial CFM50 measurement (median of 3,340 v. 2,796). These values suggest that the high saving homes are leakier and air sealing measures may generate higher savings. However, CFM50 readings are not adjusted for the size of the home and larger homes will generally have a higher CFM50. Natural air changes per hour (ACH) are calculated from CFM50 measurements using the volume of the home, number of stories and other factors. The high saving homes were generally smaller than the low savers in terms of heated area, which suggests the higher CFM50 may reflect leakier homes.

While the blower door is a valuable diagnostic tool, the blower door test is inherently inaccurate as a measurement of natural air changes per hour (ACH). Depressurizing (or pressurizing) a house with a blower door creates a set of conditions that are never replicated by nature. There are a number of reasons that conducting a comparison of the CFM50 is unlikely to yield useful results:

1. There are numerous steps that can be done incorrectly or missed by the technician. Homes need to be put into a winter state which may involve closing and latching windows and storms, opening supply registers and return grilles, covering air conditioners, etc.
2. Blower door tests are sensitive to factors such as wind conditions at the time of the test.
3. Failure to closely follow the protocols or account for conditions at the time of the test will result in less reliable estimates.

4. BPI standards allow the use of three different methods to evaluate air leakage.[6] Variations, in both the standards and their implementation from contractor to contractor, make it difficult if not impossible to analyze blower door estimates for accuracy.

House tightness for older homes may vary dramatically, and there are no standards that homes are expected to meet. In the absence of federal standards, relevant energy codes or a generally accepted range of values, the Impact Evaluation Team has no basis for assessing the accuracy of these values.

3.2 TREAT MODEL ALGORITHMS

As explained in the Methods Section, the analysis of the TREAT model algorithms was restricted to assessing possible impacts of interactive effects and whether the modeling is equally accurate for older and newer homes.

3.2.1 Interactive Effects

Interactive effects occur when measures are installed together and the impact in combination is different from the impacts of the individual measures. A simple example is installing insulation and a new heating system in the same home. As the efficiency of the heating system is higher, the savings from the insulation should be estimated at the higher efficiency level to avoid overstating savings.

To assess potential impacts of interactive effects, the savings from homes with and without specific groups of measures were compared.

- For the most part, the interactive effects calculated in TREAT are reasonable for the measures being installed.
- As expected, heating system or thermostat replacements have decreased savings when combined with other measures.
- For projects with only air sealing and insulation, there is a slight positive interactive effect, resulting in a slightly higher savings (1 to 5%) for each measure when combined. To the extent that air sealing reduces the air flow through the insulation, the performance of the

[6] ANSI/ASTM E799, CAN-CGSB 149-0010-1986 or Section 802 of the RESNET Mortgage Industry National Home Energy Rating Systems. See Standards BPI-1200-S-201x Standard Practice for Basic Analysis of Buildings p.22

insulation layer should be improved. However, it is not an effect that can be readily measured and quantified.[7]

3.2.2 Accuracy of Modeling by Age of Homes

Table 6 compares the ratio of modeled to actual consumption, average R-values and blower door readings (CFM50) for newer and older homes. This analysis clearly shows that the TREAT model overstated the annual consumption by a much wider margin for homes built before 1950. This outcome could be the result of a number of difference factors:

- Contractors may be more likely to understate the efficiency of the pre-installation conditions for older homes as compared to newer homes
- The contractors may not accurately reflect the set points in older homes, as residents may keep thermostats lower or put up with drafts or cold spots to keep heating bills lower
- The TREAT software defaults are too low on average for older homes and are selected by the contractors in lieu of entering site specific information
- The TREAT algorithms do not accurately reflect heat flow in complex, older homes

This analysis highlights the issue, but does not provide a clear way to determine the source. For example, the median R-values are lower for older homes, as would be expected; however, it is not possible to determine whether the primary issue is the inputs or the TREAT algorithms. Further investigation, possibly combining file review and site visits, would be needed to clarify the underlying causes of this wide gap in modeled annual consumption of older and newer homes.

Table 6: Modeling Accuracy and Home Age

Home Age	Count[1]	Mean Ratio of Modeled to Actual Consumption[2]	90% Confidence Interval		Median Values for Pre-install Conditions Prior to Upgrade		
			Lower	Upper	Wall R	Attic R	CFM50
Before 1950	45	2.73	2.06	3.40	6.65	6.28	3,481
1950-1969	46	1.55	1.18	1.93	8.47	7.70	2,654
1970 and later	34	1.81	1.30	2.32	10.39	10.22	2,811

[7] Lstiburek, Joseph. "WUFI*: Barking Up the Wrong Tree?" ASHRAE Journal, October, 2015. Mr. Lstiburek points out that "there is no accepted theory of combined heat and moisture flow" and one of the key missing factors is modeling airflow in complex, multilayer assemblies.

[1] This analysis included only the 125 PY2010 and PY2011 participants with complete billing data. The analysis shown in Table 3 included all participants.
[2] This column is the modeled consumption divided by the consumption from the bills averaged over houses in each age bin.

3.3 RECONCILIATION WITH UTILITY BILLS

The TREAT XML files do not explicitly state whether the modeled consumption was calibrated to the actual billing history. To investigate this issue, the Impact Evaluation Team checked two aspects of the billing reconciliation:

1. The number of utility bills entered into TREAT **-** Only 29 of the 443 projects (6.5%) had more than 12 utility bills entered.
2. The number of projects where billing data and TREAT consumption matched within 10%. Eight of the 125 projects with billing data had utility bills entered in TREAT and only one of those had TREAT consumption within 10% of the billing data consumption.

This analysis suggests that the utility billing records were rarely entered and that the presence of bills in the XML file does not indicate that the bills were used to normalize the TREAT model.[8]

The key findings from the comparison of the estimated annual consumption from TREAT to the consumption from Phase 1 billing analysis are listed below.

- The TREAT natural gas consumption was 44% higher than the billing consumption on average
- Only 7% of the TREAT models are within 10% of the billing data consumption
- 12% underestimated the annual consumption and 88% overestimated the consumption

From this analysis, it seems clear that contractors are not using the billing records to calibrate the savings in the TREAT tool. While this comparison is useful as a reality check to ensure that the savings are reasonable in the context of the actual bills, it does not identify specific modeling practices that could be contributing to the high savings. Section 3.6.1 covers the changes in modeling accuracy over the analysis period of PY2009 to PY2014.

[8] This may because of how the process of normalizing the TREAT works. To normalize the model the contractor has to manually adjust their inputs until their results more closely matches the results of the billing data. Adjusting the inputs may be complicated, since it will likely require multiple iterations of change and takes time the contractor may be unwilling to spend if they don't realize the impact on the accuracy of the results.

3.4 ERRORS IN DATA TRANSFER

When the savings from the TREAT xml files were compared to the program records, they matched almost exactly. The only errors were the occasional rounding differences of the MMBtu savings values. This analysis leads to the conclusion that discrepancies in transferring the data are not contributing to the low RRs.

3.5 INSTALLATION QUALITY

Installation quality is difficult to assess, even when site visits are conducted. In this case, the Impact Evaluation Team conducted a review of the quality assurance inspections for the projects from 2010 and 2011. While it was possible to classify the issues found on site in broad categories, there was insufficient detail to estimate an impact on savings.

A summary of the issues is shown in Table 7. The most common savings-related issues were problems with either the air sealing or insulation installation. In total, 24% of the projects had at least one problem that would reduce the energy savings and lower the realization rate. In many cases, the impact on savings appears to be minor (such as missing air sealing in part of the rim joist).

This analysis is inconclusive due to the lack of available data to investigate installation quality more thoroughly. However, it does suggest that further research is needed.

Table 7: Installation Quality Summary from 2010 and 2011 HPwES QA

Problem Category	Percent of Projects (n = 1721)[1]
Air Sealing	13%
Insulation	13%
Miscellaneous Energy	3%
Missed Opportunity	23%
Ventilation	12%
Gas Leak	9%
Combustion Test Problem	14%
Miscellaneous Non-Energy	14%
No Problems	37%

[1]This analysis was done prior to the sampling and included all of the 2010 and 2011 projects with inspections. It was not possible to conduct a similar analysis for PY2012 and PY2013 as the QA data did not contain sufficient detail.

3.6 OTHER TRENDS

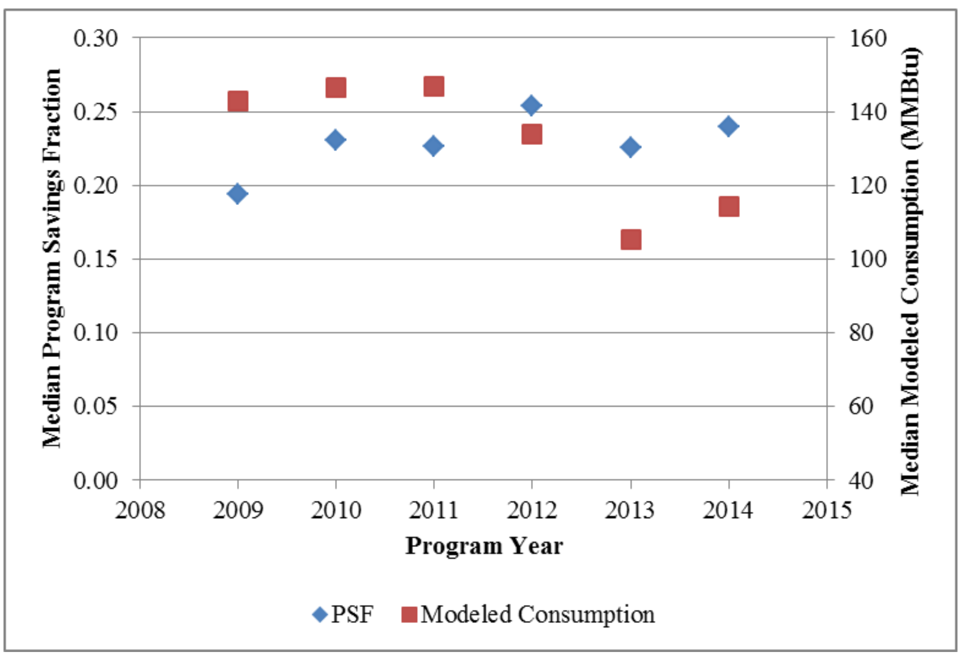
In addition to the five primary hypotheses the evaluation team investigated several other data trends that could help explain the low realization rates: trends over time, the savings investment ratio (SIR) and differences among contractors.

3.6.1 Trends over Time

In PY2011, there were a number of changes in HPwES protocols to try to improve the accuracy of the savings, including tightening up validation rules when models were uploaded to the program database and regularly collecting bills from participants. Figure 3 shows the savings fraction and modeled annual consumption by program year. This analysis was intended to assess whether these program modifications have changed the modeling practices and improved savings estimates over time. The first part of the analysis consisted of comparing the PSF and total consumption from PY2009 to PY2014.The average PSF increased slightly between 2009 and 2014. The BSF could only be calculated for two of the six years, so it was not possible to establish a trend.

In 2013 and 2014, there is a distinct drop in the median modeled annual consumption, which could be a sign that the efforts to improve the accuracy of the modeling are working. The relatively flat trend in the program savings fraction suggests that the billing savings fraction could be more reasonable in 2013 and 2014 due to the more accurate estimate of annual consumption.

Figure 3: Savings Fraction and Modeled Annual Consumption by Program Year

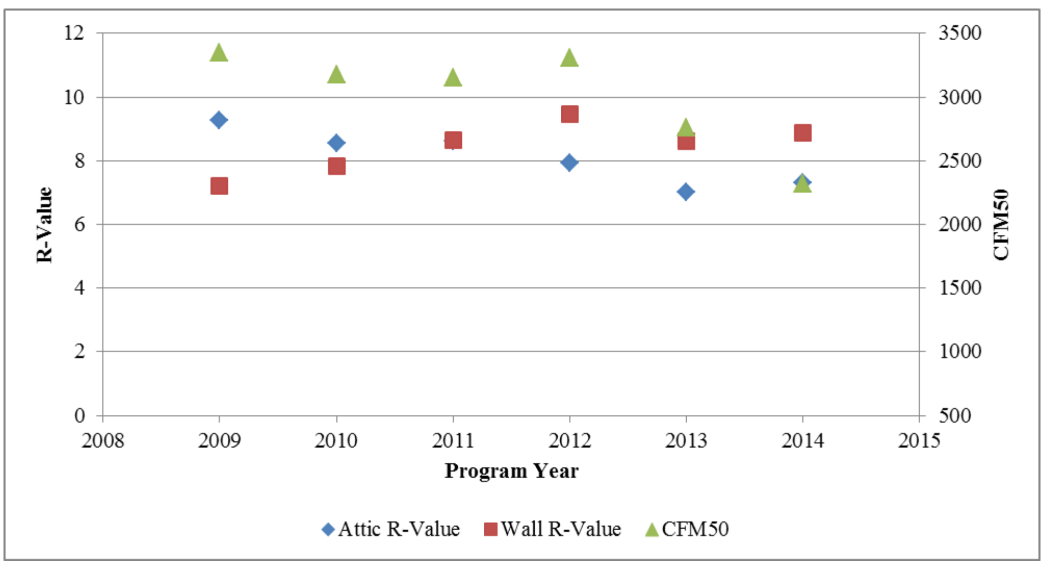


The second part of the analysis was to compare the median R-values for attic and walls and air sealing air flow (CFM) for the pre-installation conditions. Due to the changes in program protocols, we would expect to see fewer homes with very low attic and wall R-values. Between program years 2009 and 2014, this analysis shows the following:

- median wall R-value has increased slightly
- median attic R-value decreased slightly
- minimum R-values varied from year to year between R-2 and R-4 with no clear trend
- CFM50 from the blower door tests decreased slightly

Figure 4 illustrates the trends in the envelope inputs by program year. Overall, these trends seem to be somewhat contradictory. The drop in modeled annual consumption in PY2013 and PY2014 (Figure 3) suggests that modeling is improving. The CFM50 values are lower by about one-third for PY2014, which may be a contributing factor to the lower pre-installation use. As these numbers should be entered from the pre-installation blower door test, it is unclear what would be driving this decrease. In contrast, the median attic and wall R-values remain fairly consistent. While the modeled consumption dropped in PY2013 and PY2014, the median PSF remained fairly constant. This may indicate a modest improvement in estimating savings more accurately.

Figure 4: Envelope Inputs by Program Year



3.6.2 Realization Rates by Contractor

Reviewing the TREAT inputs by contractor did not show a clear pattern of low input values for any individual contractor. In addition to reviewing the individual contractor differences, several of the contractors were grouped based on high and low realization rates found in the Phase 1 analysis and compared. Some of the findings from this analysis are discussed below.

- Most contractors only had a few projects in the sample, making any conclusions uncertain. There is no indication of any particular contractors having consistently low inputs values.
- Several contractors had lower than average initial R-values, but the houses with the low R-values were all from the 1950s, so little or no insulation could be accurate.
- The high RR contractors had TREAT models with consumption values much closer to the billing data. The high RR contractors had a model consumption of 109% of the billing consumption. The low RR contractors had a model consumption of 152% of the billing consumption.

This analysis suggests the contractors with higher realization rates are making more accurate models of the original homes on average. There is no indication of any particular input driving this difference.

3.6.3 Savings to Investment Ratio

The SIR is a comparison of the monetary savings from the measures installed to the funds that would be generated by investing the cost of the measures in another financial vehicle. The interest rate of the investment can be set in the options of TREAT, and a preliminary review of the few available TREAT files suggests that contractors may be entering different values for this input. The SIR for the project is the basis for determining whether participants are eligible for financing and incentives. The Smart Energy Loan, High Efficiency Measure Incentive, Assisted Home Performance Subsidy and Residential Load Fund require a minimum SIR of 0.8 to qualify for financing.

Table 8 summarizes projects by SIR value. The low RR could occur if savings are being inflated to meet the Program criteria for financing or incentives.

Table 8: Projects by SIR Value

SIR category	Number of Projects	Percent of Projects
Less than 0.80	95	17%
0.81 to 1.20	137	24%
1.21 to 1.50	101	18%
1.51 to 2.00	84	15%

2.01 to 4.00	115	20%
Greater than 4.0	31	6%
Total	563	100%

To investigate this issue, a frequency table was constructed for different levels of SIR, as shown in Table 8. This analysis shows that 95 (17%) of the 563 projects have a SIR lower than 0.8 and 137 (24%) have SIR's between 0.8 and 1.2. Projects with just windows or AC units tend to have low SIR while projects with insulation and air sealing tend to have higher SIRs.

Few projects fall below the 0.8 SIR threshold. There are no clear differences in inputs between the bins to suggest that the input values are being manipulated to adjust the SIR. Further investigation would be needed to draw a more definitive conclusion.

3.7 CONTRACTOR SURVEY RESULTS

The Impact Evaluation Team carefully reviewed the detailed findings of the PE/MCA interviews with HPwES contractors. This section is organized by the five research topics listed in the Methods section: training, internal QA/QC, modeling and custom billing records, diagnostic testing, installation practices, use of subcontractors and comparison of HPwES and non-HPwES projects. In each section, the relevant findings from the Process Evaluation survey are listed briefly in table format, followed by a discussion of additional information that would be useful for the Impact Evaluation Team's research into the reasons for the low RRs.

3.7.1 Hiring and Training

The Process Evaluation survey included a range of questions about training and Building Performance Institute (BPI) certification, as summarized in Table 9. The table provides a summary of the Process Evaluation findings and is followed by a list of additional research areas for the Impact Evaluation Team.

Table 9: Training

Topic	Process Evaluation Survey Relevant Findings
Training	94% of surveyed contractors have some staff with energy efficiency training.
BPI Training	86% have some staff with BPI training.
Auditors/Installers	90%+ of auditors are BPI certified. 52% of contractors have no BPI certified installers.
Crew Supervisors	35% have no BPI-certified supervisors. 47% report all supervisors are BPI-certified.

Topic	Process Evaluation Survey Relevant Findings
Preference for BPI Certification	62% prefer BPI certified staff. 33% indicated BPI certified employees earn a higher wage.
HPwES Projects	60% have all of their staff work on HP projects; percentages are lower for larger firms.

Additional Research Areas for the Impact Evaluation Team

- Training Quality: What does "energy efficiency training" consist of? What is done for on-the-job training? Are specific third party workshops or seminars typically used for training?
- Costs: Who pays for training? Do contractors pay for training time?
- Access: How easy is it to sign up for training? How strongly is training encouraged? Is it an expectation?
- Types of contractors: Are there differences in training procedures between types of contractors, such as HVAC and insulation/air sealing contractors?

3.7.2 Internal Quality Assurance/Quality Control

The Process Evaluation survey covered a few questions that are related to internal QA/QC, as summarized in Table 10. The table provides a summary of the PE findings and is followed by a list of additional research areas for the Impact Evaluation Team.

Table 10: Internal Quality Assurance

Topic	Process Evaluation Survey Relevant Findings
Post Installation QA	81% of surveyed contractors do some sort of post installation QA.
Frequency of Call Backs	About 40% reported 5% of less of projects required a return visit. A quarter reported call backs for 5 to 10% of projects.

Additional Research Areas for the Impact Evaluation Team

- QA/QC Process: How are homes selected? Who performs the QA? Is it complaint-driven? Do contractors record the results? Is there a standardized list of things that get checked?
- Follow Up: What are the procedures for defining and correcting deficiencies? Are changes on site recorded in the project files?

- Diagnostic Tools: Are diagnostic tools use for troubleshooting? If so, are there specific criteria that trigger the use of these tools?

3.7.3 Modeling and Customer Billing Records

The Process Evaluation survey included an extensive set of questions about modeling and obtaining consumption data for participant projects. The key findings of that research are summarized in Table 11 and topics for further research are listed below the table.

Table 11: Modeling and Customer Billing Records

Topic	Process Evaluation Survey Relevant Findings
Modeling Software	81% of surveyed contractors use TREAT, with Real Home Analyzer at 15%.
Advantages of Modeling	90% of surveyed contractors say modeling gives their firm an advantage. Two-thirds (67%) said that being able to demonstrate savings and/or payback was a unique advantage. Some contractors mentioned modeling validates the accuracy of their audit (19%) or enhances the credibility of their recommendations (13%).
Disadvantages of Modeling	Most (85%) mentioned at least one disadvantage to modeling. About half (52%) said modeling takes too much time. Others reported the modeled results can be inaccurate (25%), the software is too complex (13%), and modeling can be a staff training burden (12%).
Consumption Data	The majority (88%) of surveyed contractors reported difficulties with acquiring the consumption data needed to calibrate the model. About one-fifth of the sample reported challenges with getting delivered fuel consumption data from either homeowners (21%) or suppliers (19%). About one-tenth (13%) noted it is difficult to calculate usage based on delivered fuel data.
Value of Audit	About one-third (33%) of surveyed contractors said they always conduct diagnostic audits. The remaining two-thirds said they are able to identify upgrade opportunities without conducting a full diagnostic audit, but many volunteered that such an approach would be inaccurate in comparison to a modeled audit.
Suggestions for Changes	Surveyed contractors suggested either changing the software used (27%) or improving/simplifying TREAT. A minority requested easier access to prior consumption data (15%). Almost one-fifth (17%) indicated no changes are needed.

Additional Research Areas for the Impact Evaluation Team

- QC Modeling Inputs: Are contractors collecting modeling data in a consistent and accurate way? Are there reasonable shortcuts that can improve the modeling results?
- Model Calibration: Do the decisions made in model calibration impact the project realization rate? Do contractors have the skills and tools to reconcile model outputs with bulk fuel deliveries?

3.7.4 Diagnostic Tools

The value and use of diagnostic tools as part of the audit was covered well in the Process Evaluation survey. Key findings are described in Table 12, followed by a list of additional research topics for the Impact Evaluation Team.

Table 12: Diagnostic Tools

Topic	Process Evaluation Survey Relevant Findings
Access to Diagnostic Equipment	100% of surveyed contractors reported their firms own blower doors, combustion analyzers and combustible gas leak detectors. Most firms (>79%) reported having access to infrared cameras, exhaust fan flow meters, pressure pan test equipment, duct blaster fan with fan speed controller, and a digital pressure and flow gauge.
Use of Diagnostic Equipment	>97% of HPwES jobs include an audit, Combustion Appliance Zone (CAZ) test, combustion efficiency, and blower door tests. Infrared inspection of insulation is conducted in 71% of HPwES jobs, followed by duct leakage testing at 36% and refrigeration diagnostics for A/C units at 32%.
Use of Diagnostic Audits	Most surveyed contractors (81%) indicated they always recommend a diagnostic audit and half reported they will not provide simple walk-through audits. Contractors who provide walk through audits indicated these cases apply only to homeowners who had an audit recently or had a single specific issue.
Perceived Value of Diagnostic Testing	Participating contractors reported conducting highly comprehensive audits and demonstrated strong support for the diagnostic audit approach to home performance.

Additional Research Areas for the Impact Evaluation Team

- Quality of Installation: Are diagnostic tools used during the installation of the efficiency measures? Does the use of diagnostic improve the quality of the outcome and increase savings?

3.7.5 Installation Practices

Exploring installation practices was not a top priority research topic in the Process Evaluation survey. The only installation practice covered in the Process Evaluation survey is the sizing of heating equipment. A summary of the findings on installation practices is presented in Table 13 below. This is an area where the Impact Evaluation Team sees the need for additional research to be able to assess potential reasons for the low RRs.

Table 13: Installation Practices

Topic	Process Evaluation Survey Relevant Findings
Heating System Sizing	Of the surveyed contractors, 66% use Manual J, 15% use TREAT, 12% use heat load calculations and 29% use other methods. (Some contractors use multiple methods.)

Additional Research Areas for the Impact Evaluation Team

The Impact Evaluation Team would like to collect more detail on installation practices. Some examples are provided below.

- Insulation: Is air sealing consistently completed prior to installing insulation? How often are insulation and air sealing combined? Are infrared cameras used to assess quality of installation? What is the standard for deciding whether installation quality is sufficient? Are blower door tests performed during installation to check progress?
- Heating systems: When is duct sealing conducted? Do contractors use the duct blaster following the installation to check results? For condensing boilers, are they checking that the temperatures are set correctly for condensing or including a modulation strategy? For furnaces, is duct balancing consistently conducted?

3.7.6 Other Issues

It is possible that other issues, such as the use of subcontractors or inconsistency of installation practices between HPwES and non-HPwES projects, could have an effect on program performance. Key findings from the Process Evaluation survey are listed in Table 14 below, followed by a brief discussion of possible avenues for future research.

Table 14: Other Relevant Issues

Topic	Process Evaluation Survey Relevant Findings
Use of Subs	Half of the contractors surveyed report sometimes using a subcontractor for HVAC installations, air sealing, insulation and/or efficient windows and doors.
HP v Non-HP: Audits and Diagnostic Tests	>97% of HPwES jobs include an audit, CAZ, combustion efficiency, and blower door tests. Infrared inspections were conducted for 71% of HPwES jobs, followed by duct leakage testing at 36%. All diagnostic tests were less common in non-HPwES jobs: 49% of all residential jobs had audits, 61% had blower door tests, 54% had infrared inspections and 24% duct leakage testing.
HP v Non-HP: Heating System Sizing	Most (87%) of surveyed contractors with experience installing ENERGY STAR HVAC measures indicated they use the same approach to sizing HVAC equipment in both HPwES and non-HPwES jobs.

Further research may consider investigating the primary areas of expertise of the contractors and how and why they utilize subcontractor arrangements. Additional information about consistency of installation practices between HPwES and non-HPwES jobs may also help to explain program performance.

SECTION 4: CONCLUSIONS

This analysis was designed to assess the reasons for the low realization rates from the previous HPwES billing analyses. While this analysis cannot be considered to be definitive as it is based on secondary research and relies on QA reports from the HPwES Program, it raises some strong possibilities. As the major sources of savings are insulation and heating system replacements, the analysis focused primarily on these measures.

Originally, five hypotheses were developed to explain the low RRs. The results of the analysis for each of the five are presented in Table 15 below. The main sources of the low RR's appear to be inaccuracy of the contractors' inputs and lack of reconciliation to billing records. This analysis suggests that contractors may be underestimating the efficiency of the home under the pre-installation conditions, particularly attic and wall R-values and heating system efficiency.

Table 15: Summary of Hypothesis Testing

Hypothesis	Evidence	Impacts	Potential Size of Impact on RR's
Contractors' inputs not accurate	Strong indication	Efficiency of pre-installation conditions may be understated, increasing both pre-install consumption and savings.	Large
TREAT model algorithms not accurate	Possible indication	Some interactive effects may slightly overstate savings; older homes may be more difficult to model accurately.	Small for natural gas
No reconciliation to bills	Strong indication	Program files rarely have bills entered; where they were entered, the modeled consumption is much higher than bills; comparison to bills provides an important reality check on savings.	Large
Errors in data transfer	No indication	Model output was compared to program tracking and the data matched.	None
Installation quality	Possible indication	Review of QA records indicated 24% of homes had some installation issue that could affect savings. From the information available, it seems that issues may be small.	Possible, needs additional research

The modeling annual consumption exceeded the actual use by 66%.[9] The incorrect R-values and heating system efficiencies appear to be a large part of this difference, as shown in Table 16, but other errors and incorrect inputs may also have an impact. For example, the modeling of secondary heating sources appears to contribute to the error.

[9] This is the median value, i.e., half the homes were below and half were above.

Finally, there are two additional issues that may be contributing to the low realization rates but need more investigation:

1. There may be some differences in modeling accuracy among contractors as the modeled consumption was substantially higher than actual consumption for the contractors with the lowest RRs. The small number of projects for each contractor makes it difficult to draw conclusions.

2. Site visits or more detailed QA reports will be needed to determine the impact of the issues with installation quality. Any differences between contractors may be resolved by improving the other issues but could be investigated further by selecting more projects from specific contractors with low or high realization rates to investigate.

Table 16: Impact of Contributing Factors to the Natural Gas Realization Rate

Contributing Factor	Percent Discrepancy between Program Reported and Evaluated Savings	
	Lower Estimate	Upper Estimate
Underestimating pre-install R-value[1]	11%	23%
Heating system efficiency[2]	4%	8%
Total Impact of Modeling Imputs[3]	18%	31%
No reconciliation to bills[4]	13%	17%
Phase 1 PY2010-2011 Realization Rate	48%	48%
Unexplained RR[5]	23%	15%

[1] The low estimate is based on the assumption that the pre-install R-value was understated by R-1.5 on average in half the homes and the upper estimate that the R-1.5 understatement was the average over all homes.
[2] The low estimate is based on the assumption that the efficiency of the existing heating system prior to replacement was understated by 2.5% on average in half the homes and the upper estimate that the 2.5% understatement was the average over all homes.
[3] There may be some overlap between the overstatement of insulation and heating system efficiency, so the totals could be slightly overstated.
[4] The low and high estimates are the 90% confidence limits for the 125 projects with billing data, multiplied by 50% to account for the overlap between reconciliation to bills and the modeling inputs. The 50% was selected by comparing the overestimate of the modeling consumption to the realization rates by BSF quartile.
[5] The portion of the RR that remains unexplained was calculating by subtracting the percent related to the reconciliation of bills from the Phase 1 RR. The modeling impacts are likely to be incorporated into the reconciliation to pre-installation bills as understating the efficiency of the existing home would have the effect of inflating the pre-installation consumption.

The analysis of trends over time suggests that the protocols added in PY2012 may be having some results in improving the modeling. A drop in the average modeled consumption per home supports this conclusion. In contrast, other modeling inputs appear to be relatively stable throughout the six years of the analysis period. There is no clear evidence that contractors are manipulating the SIR to meet the criteria for financing.

In aggregate, these findings suggest that improved modeling could have a positive effect on the RR. Starting in PY2012, the Program instituted efforts to collect participants' bills and a database tools was developed to compare savings to bills. However, it does not appear that the bills were entered into TREAT or used to validate modeling inputs.

While reconciling modeled to actual use can be a time-consuming process, it can also be a powerful tool for understanding where the inputs are incorrect, depending on how motivated the modeler is to review the inputs and consider where the errors could be introduced. Spending hours to make the modeled and actual usage agree within a small margin of error is clearly not an effective use of time. However, using the reconciliation process to review and assess the validity of the inputs could improve the modelers' understanding of how the inputs affect the results and how to improve modeling on future projects.

4.1 CONTRACTOR SURVEY DISCUSSION

The Process Evaluation survey provided a wealth of relevant and useful information. A summary of survey findings and potential areas for further research is provided in Table 17 below. Further investigation into the reasons for the low realization rates requires a level of detail that goes beyond the objectives of the Process Evaluation interviews. Combining the results of the project reviews and insights from the process surveys provides direction for future evaluation activities.

Table 17: Summary of Survey Findings and Areas for Further Research

Topic	Process Evaluation Survey Relevant Findings	Future IE Research Areas
Staff Hiring and Training	Frequency of BPI certified auditors, installers & supervisors; value of BPI training	Specifics on how training is done, what is expected, who covers the costs, etc.
Internal QA/QC	Frequency of QA/QC inspections and call backs	Content of inspections, use of diagnostic equipment and how jobs are selected for QA site visits
Modeling/ Customer Billing Records	Contractors' perceptions on the value of modeling and issues with completing the modeling and obtaining customers' bills	Accuracy of inputs and frequency of reconciliation to participants' billing records
Use of Diagnostic Testing	Frequency of use of diagnostic tools for audits, value of diagnostic audits	Use of diagnostic equipment during installations and to troubleshoot problems
Installation Practices	Heating system sizing	Wide range of other issues about installation practices

The analysis of the modeling output files indicates that a substantial part of the low RRs could be related to two aspects of modeling:

- contractors' inputs into the modeling software, particularly underestimating the efficiency of the pre-installation conditions

- lack of reconciliation to pre-install consumption (e.g., calibration)

In addition, installation quality may also be a factor and we were unable to investigate the potential impacts either through the project-level reviews or the Process Evaluation survey results. Combined, these three issues, particularly installation quality, could be affected by many of the factors initially identified for further research, including hiring and training, QA/QC procedures, modeling and use of billing records, use of diagnostic equipment and installation practices.

While the Process Evaluation survey provides valuable information regarding the extent that contractors have BPI certified staff; there is much to be learned as to the how the training is integrated into business operations and the extent that staff training impacts project outcomes. According to the Process Evaluation survey, less than half of crew supervisors are BPI certified and certification is far more prevalent in the initial audit as opposed to the staff completing the actual installations. Examining how the use of untrained installers may impact project results could provide a basis for improving program performance.

The Process Evaluation survey covered the contractors' perspectives on modeling. These responses suggest that the contractors are divided about the modeling:

- The vast majority seem to see modeling as a key component of the program and as an important way to distinguish themselves from their competitors.
- They also mention the time and training commitment as disadvantages.

This dichotomy merits further investigation to ascertain the most effective way to preserve the benefits that modeling brings while also encouraging contractors to utilize an effective modeling strategy.

The utilization of diagnostic tools is another area that warrants further study. From an impact perspective, diagnostic tools have the potential for extensive use as part of measure installation and QA/QC. Learning the extent that contractors employ techniques to integrate diagnostic tools into the installation process could provide valuable insights into how to improve program savings.

Selecting the evaluation activities is dependent on the type, quality and level of detail of the information to be obtained, as explored further in the Recommendations section.

SECTION 5: RECOMMENDATIONS

The Impact Evaluation Team understands that the HPwES Program is in a period of transition. One of the challenging aspects of the Program is using contractors as the vehicle for conducting the audits and delivering the program. Modeling can be a time consuming process and the contractors may justifiably find it frustrating.

However, understanding the modeling process and the relative impacts of specific inputs can improve their ability to estimate savings accurately. While the energy consumption in any specific home may vary due to factors that are outside of the control of the Program, improving the modeling is likely to provide estimated savings that are more in line with actual achieved savings on average over all projects.

Regardless of the delivery approach, modeling accuracy is still an important component to successful program implementation. Switching to either a more traditional third party audit or the performance contracting approach currently being used in the multifamily sector will still require reasonable estimates of the savings potential at the participating homes.

Performance contracting has not been widely adopted in the residential sector, most likely due to the small savings per home and the high uncertainty in achieving the estimated savings at any particular site. If contractors are to be expected to guarantee savings and receive a share of the money saved, they will need to be confident that the savings can be achieved, if not in a specific home, then on average over all of their participating homes.

The recommendations are divided into program and evaluation recommendations.

5.1.1 Program Recommendations

Under any scenario, understanding the modeling and how to improve the accuracy is a critical factor for program implementation. The following recommendations are related to improving the accuracy of the savings and may need to be adapted to the specific delivery mechanism.

Program Recommendation #1: Keep the Diagnostic Audit and Modeling as Core Components and Expand on Efforts to Improve Modeling

The diagnostic audit and modeling should remain core components of this program as they were clearly identified as valuable tools by both contractors and participants. Previous efforts, e.g., setting limits on the range of R-values for the pre-install conditions, appear to be having some effect. Review default assumptions for wall and attic assemblies based on the age of the home or actual wall and ceiling construction. Better documentation of combustion test results and blower door test results and confirming they match the model inputs is another step that could be taken to improve the model accuracy.

Program Recommendation #2: Improve Strategies for Accessing and Using Customer Pre-Install Billing Records

Matching modeling inputs to known pre-installation consumption patterns is likely to improve the accuracy of the modeling on average. Obtaining and using the pre-install billing records continues to be challenge, as is well known to HPwES program staff. However, reconciling the models to the bills is critical to improving the modeling results. The Impact Evaluation Team is available to discuss options.

Program Recommendation #3: Investigate the Modeling Algorithms for Electric Savings from Space Heating Measures

Over 50% of the electric savings are associated with heating-related measures. As the incidence of electric space heating is very low, the source of these savings is unclear.

5.1.2 Evaluation Recommendations

This study relied on secondary data sources and, thus, the results are suggestive rather than definitive. With that caveat, the analyses indicate that a major part of the contribution to the low RRs is underestimation of the efficiency of the home prior to the upgrades. The Impact Evaluation Team recommends the following steps for future evaluation activities.

Evaluation Recommendation #1: Consider alternative approaches to assess modeling inputs and installation quality, such as pre- and/or post-installation inspections, ride along with contractors and/or work site inspections.

5.1.3 Areas for Further Research

Through the review of the Process Evaluation survey, we have identified specific areas where additional information is needed to assess the actual impacts of the three potential contributors to the low RRs identified through the investigation into program savings. The initial impact evaluation plan included a combination of phone surveys, in-person interviews, pre-install site visits and possibly ride alongs with contractors. Another option may be to conduct site visits to verify audit inputs following the audit but before the installation of measures. Each of these strategies has strengths and weaknesses:

In Depth Phone Surveys: Phone surveys are less expensive and, thus, a larger number can be conducted, expanding the range of contractors included in the analysis. However, phone surveys generally need to be limited to half an hour to forty-five minutes, which may not be sufficient cover the variety of topics. Due to the level of detail needed, other potential evaluation activities are likely to be more productive.

In-Person Interviews: Traveling to the contractor's office to conduct the interview takes more time and resources, but may allow for a more productive interview as the interviewer can more easily develop rapport and assess body language. It also provides the opportunity to review contractors' records, if they agree to this request.

Ride Alongs: This approach allows impact evaluators to talk to the auditors and/or installers and learn about their perspective on their work, including ongoing training opportunities and internal QA. While we understand that we will most likely be assigned to the best employees and they will probably do their best work as they are being observed, the informal interactions could provide a valuable addition to other evaluation strategies.

Inspection at Time of Installation or Post-Installation: The purpose of these site inspections would be to assess the quality of the installation. The approach would depend on the nature and timing of the installation. For instance, observing the installation of wall cavity insulation would be more useful than a post-installation inspection. Other measures, such as heating system replacement, can be easily verified through traditional post-installation inspections. For measures that require inspection prior to completion of the project, one approach would be for the evaluator to tag along on existing QA/QC visits. This approach maintains third party independence while minimizing the number of visits to participant's homes.

Pre-Install Verification Site Visits: Conducting site visits between the audit and the installation could allow us to verify the inputs into the model. This approach would allow evaluators to confirm or disprove the findings from our previous review of model outputs suggesting that contractors may be underestimating the efficiency of the pre-installation conditions.

Phone surveys are less likely to be an effective tool as the topics are highly detailed. Future impact evaluation activities should consider the four evaluation activities as shown in Table 18.

Table 18: Summary of Future Research Areas and Evaluation Activities

Topic	In-Person Interview	Ride Alongs	Pre-Install Site Visit	Post-Install Site Visit
Hiring and Training	√	√		
Internal QA/QC	√	√		
Modeling/Customer Billing Records	√		√	
Use of Diagnostic Testing	√	√	√	
Installation Practices	√	√		√