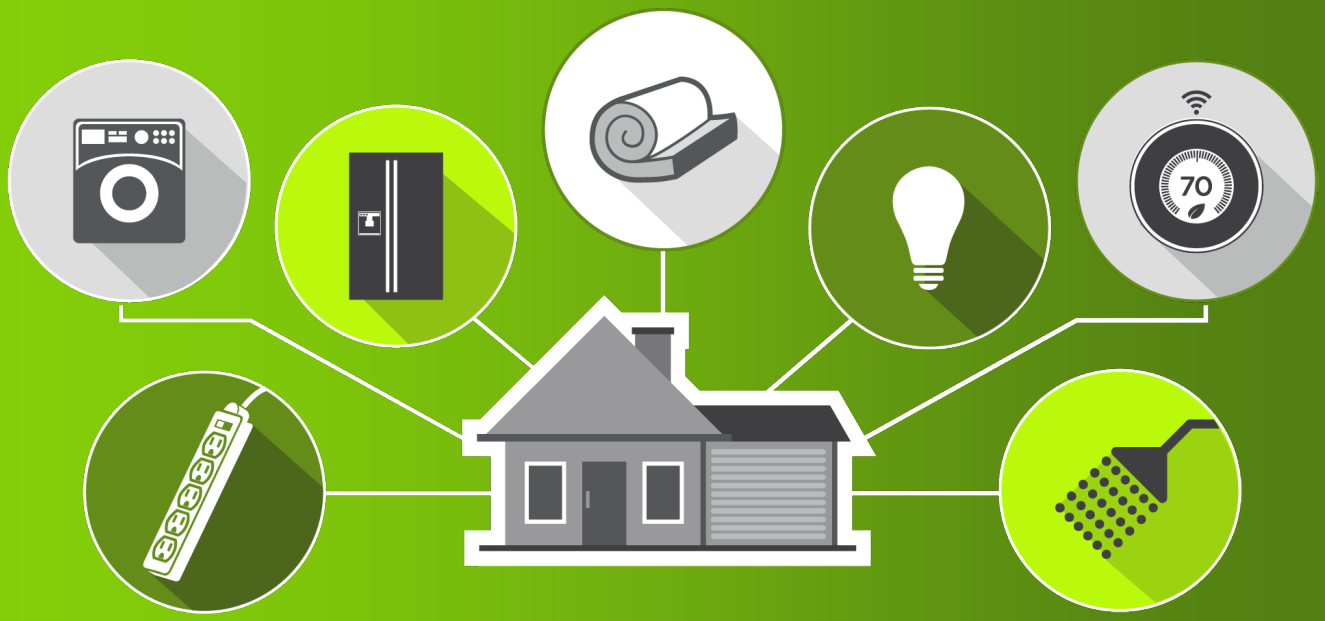


HOME ENERGY SERVICES

Realization Rate Assessment (RES 39)

MARCH 2020

PRODUCED IN COLLABORATION BETWEEN NAVIGANT cadeo

PREPARED FOR: THE ELECTRIC AND GAS PROGRAM ADMINISTRATORS OF MASSACHUSETTS
PART OF THE RESIDENTIAL EVALUATION PROGRAM AREA

HES: REALIZATION RATE ASSESSMENT

ABOUT

The previous Home Energy Services (HES) impact evaluation (RES 34) yielded ex post realization rates for air sealing and insulation (collectively referred to as weatherization) measures installed through the Program Administrators' (PAs) largest residential program. Statewide, RES 34 found that the average natural gas-heated HES participant receiving air sealing and/or insulation realized 73% of the total energy savings predicted by the HES Lead Vendor (LV) or Home Performance Contractor (HPC) that treated their home. The results of RES 34 were nearly identical to the realization rate (0.76) determined by the prior HES impact evaluation (completed in 2013) for natural gas weatherization. The similarity of these results, as well as the fact that none of the PAs or HES LVs achieved a statistically significant realization rate for natural gas higher than 0.85 as part of either evaluation, indicates a persistent disconnect between LV ex ante and evaluator ex post savings.

RESEARCH QUESTIONS

This study sought answers to the following two key research questions:

1 | **Why are the evaluation realization rates for weatherization measure not closer to 1.0?**

2 | **What changes would improve vendor savings estimates?**

To answer these overarching research questions, the team assessed each the following supporting questions, which could be contributing the persistent realization rate disparity:

1. Methodological Differences. Do methodological differences between the approaches used to estimate ex ante and ex post savings contribute to the observed disparity?

2. Differences by Participant Type. Is the disconnect between ex ante and ex post more pronounced for certain subsets of HES participants (e.g., single family detached vs. 2-4 units or participants that received different types of insulation)?

3. Model and Process Integrity. Are the tools and processes that LVs use robust? Do the LVs models rely on the best available information? To what extent, if any, does the experience and approach used by the Energy Specialist (ES) or Home Performance Contractor (HPC) impact modeling accuracy?

4. Customization. How does each LV customize their savings estimates to reflect a specific participating customer?

5. Behavior Change. Are participants changing their behavior after being weatherized in a way that contributes to the disparity?

METHODOLOGY

There are many factors that potentially contribute to a difference between ex ante and ex post savings. These include, but are not limited to:

- Mischaracterization of existing building conditions (e.g., inaccurate building envelope assumptions and/or equipment efficiencies)
- Modeling issues (e.g., incorrect weather assumptions, inaccurate engineering algorithms, and/or misapplication of interactive effects)
- Rebound effect (e.g., participants maintaining higher heating setpoints after weatherization)
- Fundamental differences in estimation methodologies (e.g., customer-specific building simulations versus aggregated evaluation billing analysis)

To understand how these factors are impacting realization rates and answer the two overarching research questions above, the evaluation team undertook the following four, complementary tasks:

Task 1 | Engineering Review of LV Modeling Software

Our team began the study by reviewing the modeling tools used by HES LVs. The review included an independent assessment of the model itself (i.e., reviewing model inputs, data sources, and model functionality), as well as in-depth interviews with each LV's modeling teams.

Task 2 | Interviews with Assessors

Following our interviews with the LV modeling teams, we interviewed 20 Energy Specialists (n=14) and HPCs (n=6) that, collectively, estimate weatherization savings using every LV tool. The interviews into how assessors use the LV tools "in the field" (i.e., the inputs they update and ones they do not).

Task 3 | Survey Participants

To assess the extent to which HES participants (i.e. rebound effect or snapback), our team surveyed 321 HES participants that weatherized their home in 2018. These surveys, completed online, includes participants from all seven PAs and focused on behavioral changes to participant's heating and cooling practices following weatherization.

Task 4 | Compare Modeled Consumption and Billing Data

To understand how participant's LV modeled consumption compared to their actual consumption (i.e., PA billing data), our team compared each participant's actual pre- (2016) and post- (2018) HES annual consumption with modeled annual estimates provided by the HES LVs.

KEY FINDINGS

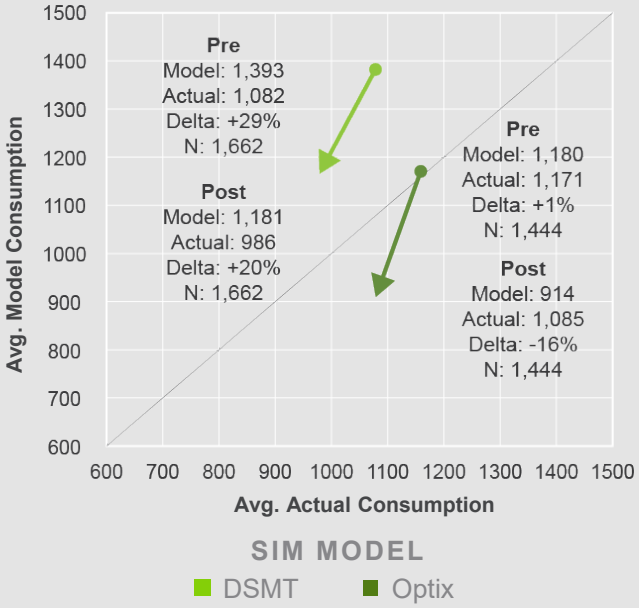
1. Methodological Differences

Key Finding: Independent of impact evaluation regression model, a direct comparison of modeled and actual consumption reveals that both whole-home simulation models overpredict total consumption and/or savings.

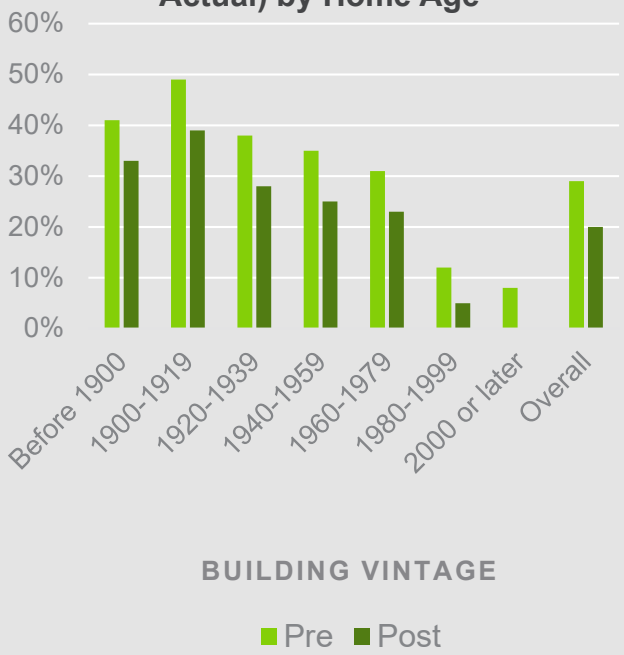
2. Differences by Participation Type

Key Finding: Both whole-home simulations less accurately estimated the consumption of older homes, which typically receive more insulation than newer homes.

Average Pre- and Post-Consumption Estimates by Software



Average Difference (Modeled vs. Actual) by Home Age



3. Model and Process Integrity

Key Finding: All three LVs use robust models that rely on MA-specific data. There does not appear to be meaningful differences in how ESs or HPCs use these tools in the field; most shared a similar data collection and modeling approach and reported they infrequently change model default values.

4. Model Customization

Key Finding: Energy Specialists and HPCs do not typically utilize customer-specific billing data when modeling a participant's home and estimating weatherization savings. However, assessors typically have access to participant's usage information (i.e., the participant shares a copy of their bill). In these instances, some assessors will use the bill to do a "sanity check" on the model's estimated savings.

5. Participant Behavior Change

Key Finding: Rebound effect is not a meaningful contributor to the realization rate disparity observed in the previous two HES impact evaluations

RECOMMENDATIONS

Based on the key findings above, our team developed recommendations for the PAs as they look to answer the study's second core question: **What changes would improve vendor savings estimates?**

RECOMMENDATION 1: Integrate Participant Consumption Data

Our model reviews and LV interviews determined that none of the HES LVs currently use a participant's pre-program energy consumption data to model their potential weatherization savings in a formal or systematic way. Not using participant-specific consumption data means the weatherization savings modeled for each participant is disconnected from that participant's actual energy usage. Our team believes this fundamental disconnect is likely a significant contributor to observed disparity in historical weatherization realization rates.

Since improving realization rates requires the HES LV models to accurately characterize energy consumption pre- and post-intervention, we recommend that PAs and LVs work together to formally integrate participant's actual historical energy consumption into the weatherization modeling process. We recognize that PAs and LVs have different legacy data management infrastructures and policies and, as such, also recognize the importance of providing each PA and LV with room for operationalizing this consideration.

Ideally, the LVs can load participant's consumption data into their models in an automated way prior to each assessment. This would prevent assessors from relying on participants to share a copy of their bill or from manually retrieving consumption data. However, if automated integration of consumption data into the modeling software is not technically possible—or requires significant time to implement—the PAs and LVs should take simpler, near-term steps toward integrating participant's consumption history. As an example, PAs could ensure that assessors have access to participant's consumption records and require them to review that information prior to the assessment.

RECOMMENDATION 2: Formalize Sanity Checks

Our assessor interviews also determined that some assessors perform an informal "sanity check" to confirm that their modeled total consumption and savings values are reasonable using a copy of the participant's bill. However, this practice is assessor-specific and only occurs when a participant proactively provides a copy of their bill or can provide one upon request.

RECOMMENDATIONS

Consistent with the concept of better leveraging participant-specific consumption data to improve model accuracy, we recommend formalizing these sanity checks. There is variation in the capabilities of PA billing systems and LVs modeling systems, thus the sanity check could happen through either an automated or manual process. As an example, it may be possible to add an automated alert to the modeling software itself notifies the assessor if modeled consumption is beyond +/-10% of the home's actual consumption or the modeled weatherization savings exceed 35% of a participant's actual pre-program consumption. (Our team has offered placeholder values for these alert thresholds, but we suggest the PAs and LVs work together to select the appropriate values.) Alternatively, an example of a manual sanity check is a program requirement for the assessor to confirm that the model's consumption and/or savings fall within the designated range before the LV software produces a final assessment (and, if it does not, document

Whether implemented using an automated or manual approach, the formalization of these sanity checks will require all assessors to adopt this best practice and mitigate outlier savings estimates.

RECOMMENDATION 3: Revisit Existing R-Value Assumptions

Our team found meaningful differences between modeled and actual consumption for Eversource and National Grid participants across all types of insulation, particularly in older homes that received multiple types of insulation. (Our team was unable to complete a similar analysis for the other PAs.) Through the team's literature review and an interview with a building simulation industry expert, the team found anecdotal evidence that under-characterizing the R-value of the uninsulated building shell (especially in older homes) can be a source of overestimation of heating system energy use. Our team attempted to assess the default R-values associated with air films and the non-insulation portions of the building shell in each model (including wall assemblies, e.g. sheathing, siding, drywall), but we did not have access to this specific information. This and other background assumptions and/or engineering calculations were unavailable behind the model's user interface. For this reason, we recommend the LVs closely review all default R-value assumptions and consider using, at a minimum, a default of R-3 to reflect the thermal properties of a completely uninsulated wall.

DEVELOPED BY NAVIGANT cadeo

CONTRIBUTORS TO THIS REPORT:
Doug Bruchs, Fred Schaefer, Bretnie Eschenbach, and Jonah Hessels, Cadeo Group;
David Basak and Ken Seiden, Navigant Consulting.

BERKSHIRE GAS

BLACKSTONE GAS COMPANY

Cape Light Compact

Columbia Gas of Massachusetts
A NiSource Company

EVERSOURCE ENERGY

Liberty Utilities

mass save

nationalgrid

Unitil

Home Energy Services Realization Rate Assessment (RES 39)

Final Report

Prepared for:

The Electric and Gas Program Administrators of Massachusetts
Part of the Residential Evaluation Program Area

Submitted by:
Navigant, A Guidehouse Company
1375 Walnut Street
Suite 100
Boulder, CO 80302

303.728.2500
guidehouse.com

Reference No.: 209974
March 19, 2020

TABLE OF CONTENTS

Executive Summary iii
1. INTRODUCTION 1
1.1 Research Goals 1
1.2 Research Questions 2
1.3 About HES 2
1.3.1 Other Fuels 3
2. METHODOLOGY 5
3. KEY FINDINGS 7
3.1 Methodological Differences 7
3.2 Differences by Participant Type 10
3.3 Model and Process Integrity 2
3.4 Model Customization 1
3.5 Participant Behavior Change 2
4. RECOMMENDATIONS 5
Appendix A. COMPARATIVE ANALYSIS - ELECTRIC RESULTS A-1

LIST OF FIGURES

Figure 1. Comparison of Pre- and Post-Modeled and Actual Natural Gas Consumption: 2017 Eversource Weatherization Participants using DSMT (n=1,662) 8
Figure 2. Comparison of Pre- and Post-Modeled and Actual Natural Gas Consumption: 2017 National Grid Weatherization Participants using Optix (n=1,444) 9
Figure 3. Average Pre- and Post-Consumption Estimates by Software 10
Figure 4. Comparison of Pre- and Post-Modeled and Actual Natural Gas Consumption: For Subsets of 2017 Eversource Weatherization Participants (n=1,662) 1
Figure 5. Comparison of Pre- and Post-Modeled and Actual Natural Gas Consumption: For Subsets of 2017 National Grid Weatherization Participants (n=1,444) 2
Figure 6. Optix Duct & Distribution Module 3
Figure 7. Self-Report Changes to Heating and Cooling Temperature Set Points After Weatherization through HES 3
Figure 8. Comparison of Pre- and Post-Modeled and Actual Electric Consumption: 2017 Eversource Weatherization Participants A-1
Figure 9 Comparison of Pre- and Post-Modeled and Actual Electric Consumption: 2017 National Grid Weatherization Participants A-2
Figure 10. Average Pre- and Post-Consumption Estimates by Software A-3
Figure 11. Comparison of Pre- and Post-Modeled and Actual Electric Consumption: For Subsets of 2017 Eversource Weatherization Participants (n=349) A-4
Figure 12. Comparison of Pre- and Post-Modeled and Actual Electric Consumption: For Subsets of 2017 National Grid Weatherization Participants (n=165) A-5

1. INTRODUCTION

EX ANTE vs. EX POST

In the context of this report, the Latin term ex ante (which means "from before") refers to the energy savings predicted for a given customer by an HES LV using their proprietary building simulation software prior to their participation in HES. Conversely, ex post refers to the same savings as estimated by an independent evaluator after the customer's participation in the program using an econometric billing analysis.

The previous Home Energy Services (HES)[1] impact evaluation (RES 34) yielded ex post realization rates for air sealing and insulation (collectively referred to as weatherization) measures installed by the Program Administrators' (PAs).[2] Statewide, RES 34 found that the average natural gas-heated HES participant receiving air sealing and/or insulation realized 73%[3] of the total energy savings predicted by the HES Lead Vendor (LV) or Home Performance Contractor (HPC) that treated their home.

The results of RES 34 were nearly identical to the realization rate (76%) determined by the prior HES impact evaluation (completed in 2013) for natural gas weatherization.[4] The similarity of these results, as well as the fact that none of the PAs or HES LVs achieved a statistically significant realization rate for natural gas higher than 85% as part of either evaluation, indicates a persistent disconnect between LV ex ante and evaluator ex post savings.

Lower realization rates are problematic and therefore worthy of further evaluation research because the PAs use LV ex ante estimates to report savings and track progress toward their annual and three-year savings goals. While the PAs typically use the most recent realization rates to anticipate future evaluation adjustments, uncertainty around that magnitude of the adjustment makes it more difficult for implementers to feel confident about their progress toward goals.

The Massachusetts Residential Retrofit Research Area team, led by Navigant and joined by Cadeo and ILLUME Advising (referred to as "the evaluation team" throughout this document) completed this study on behalf of the Massachusetts PAs with oversight from the Energy Efficiency Advisory Council (EEAC) consultants. Cadeo led this specific study, including all primary data collection, analysis, and reporting.

1.1 Research Goals

The goal of this study was to investigate potential causes for the disparity between ex ante and ex post weatherization savings apparent in the last two evaluations and to help the PAs (and their LVs) close this persistent savings gap by recommending specific actions that the PAs and/or LVs can take that might minimize the difference in future evaluations.

[1] As part of the PA's 2019-2021 Three-Year Energy Efficiency Plan, the PAs have begun transitioning HES into the Residential Coordinated Delivery (RCD) program. RCD, which combines the HES and the PA's legacy multifamily program, will serve a single, streamlined program for all non-new construction and non-low-income residential customers. Since the transition to RCD is in progress and because this report focuses relies on historical HES data, the team refers to HES, not RCD, throughout this report. Regardless of the terminology used, the results of this study are directly applicable to RCD prospectively: the Lead Vendors implementing RCD as the same Lead Vendors—using the same modeling tools—that implemented HES.

[2] http://ma-eeac.org/wordpress/wp-content/uploads/RES34_HES-Impact-Evaluation-Report-with-ES_FINAL_29AUG2018.pdf

[3] PA-specific results varied from 68% to 85% for natural gas.

[4] http://ma-eeac.org/wordpress/wp-content/uploads/Home-Energy-Services-Realization-Rate-Results-Memo-6-28-12.pdf

1.2 Research Questions

This study sought answers to the following two key research questions:

1. Why are the evaluation realization rates for weatherization measure not closer to 1.0?[5]
2. What changes would improve vendor savings estimates?

To answer these overarching research questions, the team assessed each the following supporting questions, which could be contributing the persistent realization rate disparity:

1. **Methodological Differences**. Do methodological differences between the approaches used to estimate ex ante and ex post savings contribute to the observed disparity??
2. **Differences by Participant Type.** Is the disconnect between ex ante and ex post more pronounced for certain subsets of HES participants (e.g., single family detached vs. 2-4 units or participants that received different types of insulation)?
3. **Model and Process Integrity.** Are the tools and processes that LVs use robust? Do the LVs models rely on the best available information? To what extent, if any, does the experience and approach used by the Energy Specialist (ES) or Home Performance Contractor (HPC) impact modeling accuracy?
4. **Customization.** How does each LV customize their savings estimates to reflect a specific participating customer?
5. **Behavior Change**. Are participants changing their behavior after being weatherized in a way that contributes to the disparity?

1.3 About HES

Before discussing the study's methodology or its key findings, it is important to first understand how the PAs deliver HES. Statewide, the seven PAs rely on four different LVs to implement HES: CLEAResult, RISE Engineering (RISE), the Center for EcoTechnology (CET), and Abode. In turn, each LV uses a different building modeling software to estimate weatherization savings for the HES participants they assess. In the case of CLEAResult, they use two different building simulation tools depending on whether the participant is a National Grid or Eversource customer. RISE, similarly, uses two tools – one for National Grid customers living on the Cape and another for all the other PA customers they serve.[6]

Table 1-1 summarizes the relationships between PAs and LVs, as well as the applicable building simulation software for each combination.

[5] Realization rates are only applicable to weatherization (i.e., insulation and air sealing) because the PAs do not use LV-generated savings estimates for any other HES measures. The PAs rely on deemed savings, codified in their TRM, for all other HES measures.

[6] When multiple PAs serve a customer, HES associates that customer with the PA that provides their heating fuel. For example, take a customer with a natural gas furnace received their natural gas from Columbia Gas and their electricity from Eversource. In the context of HES, the initiative would designate such a customer as a Columbia Gas participant and RISE, Columbia Gas' LV, would be responsible for the customer's home energy assessment. Because the PAs' coordinate closely, RISE would leverage their access to the customer's home and install electricity-saving direct install measures during the home energy assessment. However, Eversource, the customer's electric utility, would pay for those measures and claim the associated electric savings.

Table 1-1. PA, LV, and Model Mapping

PA	LV	Simulation Model	Model Type
National Grid	CLEAResult, RISE (On Cape) Abode	Optix*	Whole-home Simulation
Eversource	CLEAResult	Demand Side Management Tracker (DSMT)**	Whole-home Simulation
Berkshire Gas	CET	CET	Weatherization-specific Algorithm
Columbia Gas	RISE	EPLUS	Weatherization-specific Algorithm
Cape Light Compact			
Liberty Gas			
Unitil			

*Optix is a building simulation tool created by EnergySavvy. National Grid had CLEAResult transition from DSMT to Optix starting in November 2017 to align HES data with other EnergySavvy program-delivery tools that National Grid was using.

**CLEAResult transitioned their Energy Specialists to DSMT in August 2018. Previously, CLEAResult's Energy Specialists used their proprietary building simulation software was called EMHome. (HPCs in Eversource's service territory still use EMHome.) According to interviewed CLEAResult staff, the transition to DSMT improved the tool's functionality but did not impact any of the underlying engineering calculations. Consequently, the transition from EMHome to DSMT does not have any bearing on the results of this study.

Table 1-1 also identifies each building simulation software as one of two types: a whole-home simulation or weatherization-specific algorithm. The names of these model types closely describe how they function. The differences between model types had implications for the study. Some of the evaluation activities described in the next section were only applicable for a certain model type. For example, the Task 4 Comparative Analysis was only possible for whole-home simulations. As such we note, throughout the report, if a specific activity or resulting finding is only relevant for a subset of the PAs, LVs, or model types.

MORE ABOUT MODEL TYPES

A whole-home simulation models an entire home and simulates the energy consumption for that home in a typical weather year. These models estimate savings by calculating the difference in simulated annual energy consumption for the entire home before and after weatherization. Alternatively, a weatherization-specific algorithm estimates savings by collecting a smaller amount of information related to the weatherization (e.g. pre/post R values, and square feet of added insulation, heating system efficiency, etc.) which assessors plug into an equation to estimate energy savings. Unlike the whole-home simulation, the algorithmic approach does not produce household level energy consumption. While the building simulation models incorporate more explicit factors in their savings calculations (e.g. interactive effects, duct and distribution losses, home types), both engineering algorithms employ a standardized "correction factor" to account for implicit factors that impact weatherization savings.

1.3.1 Other Fuels

Roughly half of the households that install weatherization through the HES program are natural gas-heated. As described above, the impetus for this study was the need to explore the potential causes of consistently low realization rates associated with those weatherization participants.

However, the findings and recommendations of this study are not limited to natural gas participants and pertain to all HES weatherization participants. It is important to note that our team's model reviews and discussions with LVs, Energy Specialists, and HPCs related to the program's overall modeling process and were not specific to participants that heat with natural gas. Similarly, our participant surveys included customers relying on all heating sources.

The lone exception is the Task 4 comparative analysis, which is described in more detail in the following section. Because that task consisted of comparing LV modeled consumption with participant-specific historical consumption, our analysis was limited to the natural gas and electric consumption data provided by the PAs. However, it is likely the results of that analysis are also generally applicable to delivered fuel heated HES participants (~40%) since LVs model weatherization savings for those participants using the same modeling software.

2. METHODOLOGY

There are many factors that potentially contribute to a difference between ex ante and ex post savings. These include, but are not limited to:

- Mischaracterization of existing building conditions (e.g., inaccurate building envelope assumptions and/or equipment efficiencies)
- Modeling issues (e.g., incorrect weather assumptions, inaccurate engineering algorithms, and/or misapplication of interactive effects)
- Rebound effect (e.g., participants maintaining higher heating setpoints after weatherization)
- Fundamental differences in estimation methodologies (e.g., customer-specific building simulations versus aggregated evaluation billing analysis)

To understand how these factors are impacting realization rates and answer the two overarching research questions above, the evaluation team undertook the following four, complementary tasks.

Task 1 | Engineering Review of LV Modeling Software

Our team began the study by reviewing the modeling tools used by HES LVs. The review included an independent assessment of the model itself (i.e., reviewing model inputs, data sources, and model functionality), as well as in-depth interviews with each LV's modeling teams. We used the interview to build upon our initial review and identify any potential systemic causes of savings overestimation. Given the technical nature of the discussion, the team summarized its discussion and confirmed the key takeaways with the interviewee following each interview. Our contact with the LV modeling teams did not end with these interviews; our team exchanged e-mails and phone calls throughout the study.

Task 2 | Interviews with Assessors

Following our interviews with the LV modeling teams, we interviewed 20 Energy Specialists (n=14) and HPCs (n=6) that, collectively, estimate weatherization savings using every LV tool.[7] The interviews complemented the top-down perspective offered by the LV modeling teams and provided our team with insight into how assessors use the tools "in the field" (i.e., the inputs they update and ones they do not).

Task 3 | Survey Participants

Rebound effect (i.e., participants changing their behavior after weatherizing their home in a way that diminishes savings) is another potential contributor to the observed disparity between ex ante and ex post savings estimates. To assess the extent to which recent HES participants took actions that produce a rebound effect, our team surveyed 321 HES participants that weatherized their home in 2018. These surveys, completed online, includes participants from all seven PAs and focused on behavioral changes to participant's heating and cooling practices following weatherization.

Task 4 | Compare Modeled Consumption and Billing Data

To understand how participant's LV modeled consumption compared to their actual consumption (i.e., PA billing data), our team analyzed LV and PA data for over 3,000 2017 HES participants. Our team

[7] Energy Specialists are home energy assessors employed by the HES LVs that are typically assigned to customers who sign up for HES without specifying a preference to work with a certain Home Energy Performance (HPC) contractor. The Energy Specialists' role is limited to the assessment itself; if the customer decides to move ahead with an Energy Specialist's recommendation, the installation is typically completed by a program-approved independent installation contractors (IICs). Conversely, HPCs are independent contractors approved by the PAs to complete the home energy assessment, as well as install any recommended efficiency upgrades that result from the assessment. HPCs often market HES on their own. For a more detailed explanation of the difference between Energy Specialists and HPCs and the PAs' overall delivery of HES, see the previous HES Process Evaluation (March 2018).

compared each participant's actual pre- (2016) and post- (2018) HES annual consumption with modeled annual estimates provided by the HES LVs. In addition to looking at overall trends, our team also investigated whether the relationship between actual and modeled consumption differed for specific participant and/or building types (e.g., home size, type, and size, as well the type of insulation the participant installed and their heating fuel).

3. KEY FINDINGS

The evaluation tasks described above provided our team with two different types of information that we used to inform the key findings in this section. The participant survey (Task 3) and comparative analysis (Task 4) yielded quantitative findings (about participant's post-program behavior and how LVs' modeled energy consumption compared to the participant's actual energy consumption, respectively). Conversely, our in-depth interviews with LVs (Task 1) and with energy specialists and HPCs (Task 2) provided qualitative insights into how the tools themselves, as well as how HES stakeholders deploy them in the field.

We used the collective quantitative and qualitative information our team gathered to investigate each of the five supporting research questions listed above to address the study's first of two core questions: **Why are the evaluation realization rates for weatherization measure not closer to 1.0?**

EVOLVING PROGRAMS

It is important to note that the PAs and their LVs—in their ongoing effort to maximize savings and improve the participant experience—regularly updates how they deliver HES. This includes how they model energy consumption and estimate energy savings from weatherization measures.

Readers of this report should understand that the findings in this section reflect the program's weatherization model practices at the time each evaluation activity occurred. This means that the interview-based findings reflect LV and Energy Specialist/HPC modeling practices as of May 2019. Similarly, the comparative analysis relied on energy bills from 2016 through 2018 for customers that participated in HES in 2017. The findings of that analysis therefore reflect the relationship between modeled and actual consumption for participants at that time. Lastly, the participant surveys that our team completed were with a sample of HES participants that weatherized their home in 2018.

In light of their continuous effort to evolve HES, it is important to acknowledge that the PAs and LVs may have since made changes that impact the findings summarized in this section.

3.1 Methodological Differences

Relevant Research Question: Do methodological differences between the approaches used to estimate ex ante and ex post contribute to the observed disparity?

Key Finding: Independent of impact evaluation regression model, a direct comparison of modeled and actual consumption reveals that **both whole-home simulation models overpredict total consumption and/or savings.**

As previously noted, the approaches that LVs and evaluators take to estimate weatherization savings are fundamentally different. LVs estimate ex ante savings for each participant using customized, participant-specific building simulation models, whereas evaluators estimate ex post savings based on pooled econometric models that rely on PA billing data for hundreds of participating and nonparticipating customers.[8] Consequently, it is possible that the realization rate disparity apparent in the last two HES impact evaluations is, at least in part, a function of these different methodologies.

To assess this possibility, our team used the Comparative Analysis to directly compare LV's pre-HES and post-HES **modeled** annual consumption estimates for 3,106 weatherized natural gas heated participants

[8] This statement is true for weatherization savings in homes where the primary heating fuel is natural gas or electricity (i.e., the PAs can provide billing data for participants). For homes heated by delivered fuels (i.e., when the PAs cannot provide billing data), previous HES impact evaluations have also relied on building simulation models, albeit aggregated models (not participant-specific models) that reflect the average condition in delivered fuel homes.

with the PA's **actual** annual consumption data for those same participants.[9] Since the PA's billing data reflects all natural gas usage in the home, which may include natural gas water heating or cooking end-uses, the analysis was only relevant for Eversource and National Grid participants modeled using CLEAResult and EnergySavvy's whole-home simulation software.

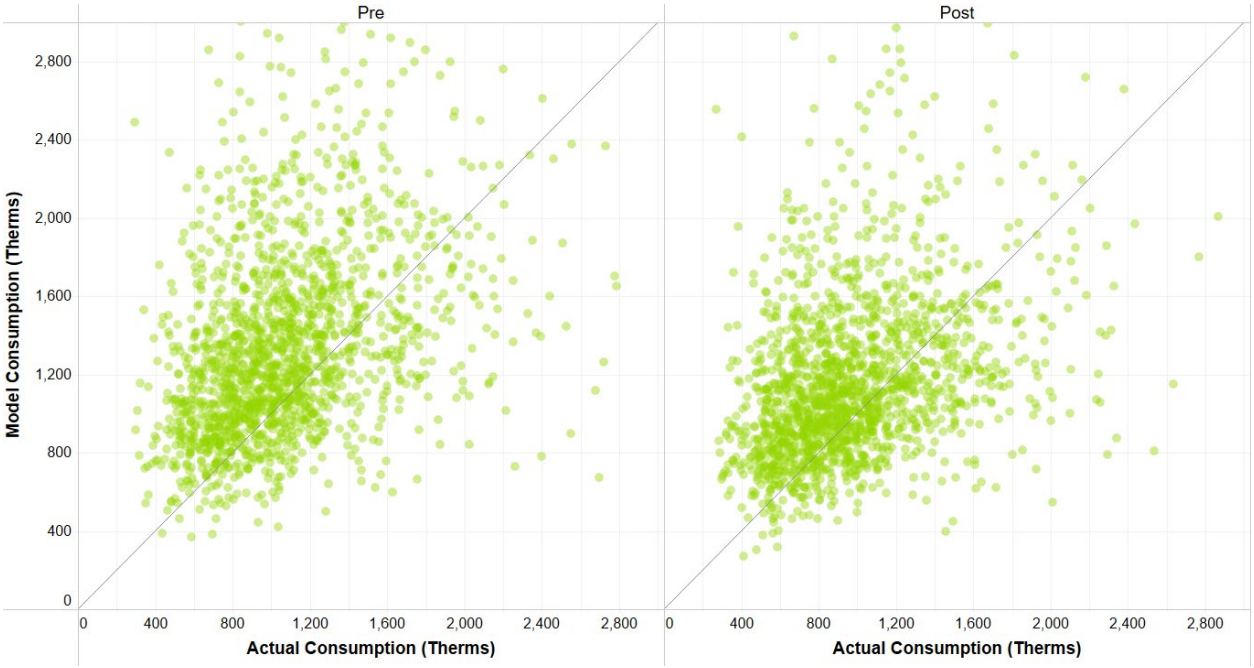
Figure 1 and Figure 2 compare modeled and actual annual energy consumption before (pre, on the left) and after (post, on the right) installing weatherization through HES for 2017 Eversource and National Grid participants. These scatterplots include a dot for each weatherized natural gas HES participant and indicate how that participant's modeled consumption compared to their actual consumption. Similar results are provided for customers in electrically heated home (~5% of all HES participants) in the report's appendix.

INTERPRETING SCATTERPLOTS

A dot anywhere along on the diagonal line through the middle of the figure indicates that participant's modeled and actual consumption were the **same**. Dots above the diagonal line indicate that the LVs' modeled consumption was **greater** than the participant's actual consumption, whereas dots below the line indicate modeled consumption was **less** than actual consumption.

As evident in these figures, CLEAResult's DSMT model generally overpredicted annual energy consumption in both the pre- and post-period. EnergySavvy's Optix model, which CLEAResult and RISE use to model National Grid participants, was quite accurate in the pre-period, but underestimated consumption (thereby overestimating savings) in the post-period.

Figure 1. Comparison of Pre- and Post-Modeled and Actual Natural Gas Consumption: 2017 Eversource Weatherization Participants using DSMT (n=1,662)



[9] The team undertook a similar analysis for electrically heated HES weatherization participants. The results for electrically heated participants are provided in 4.Appendix A.

Figure 2. Comparison of Pre- and Post-Modeled and Actual Natural Gas Consumption: 2017 National Grid Weatherization Participants using Optix (n=1,444)

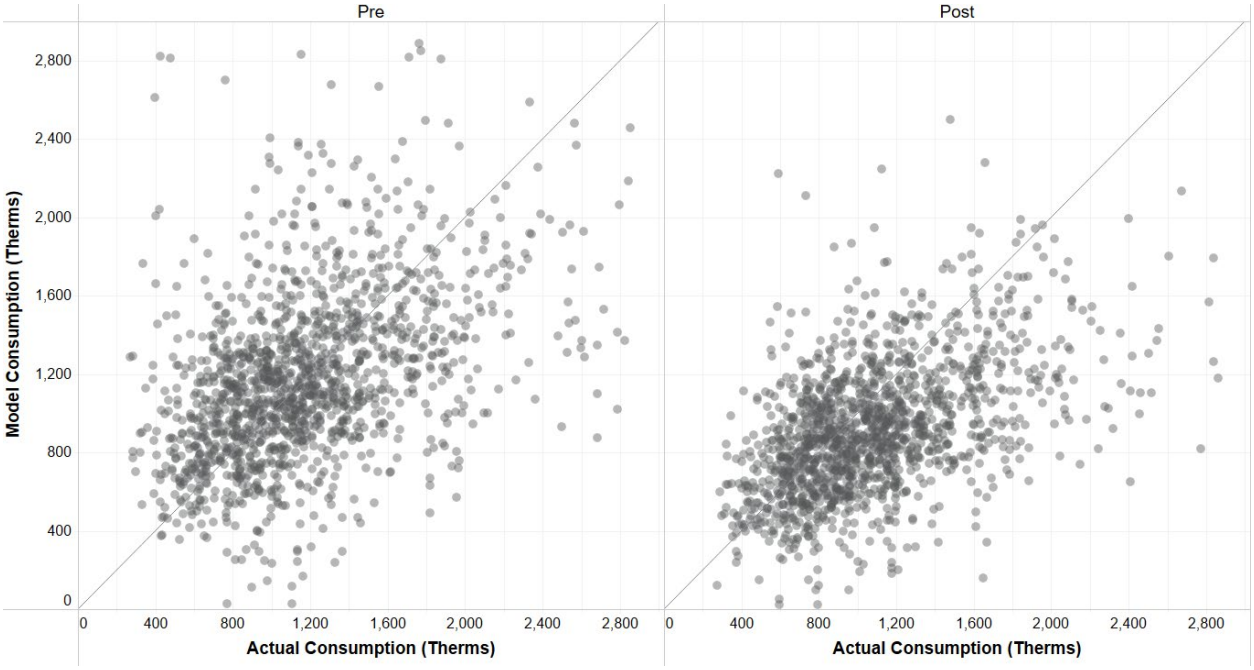


Figure 3 offers a different perspective on the same data. Rather than a dot for each participant, Figure 3 includes only four points representing the average pre- and post-value for Eversource and National Grid participants.

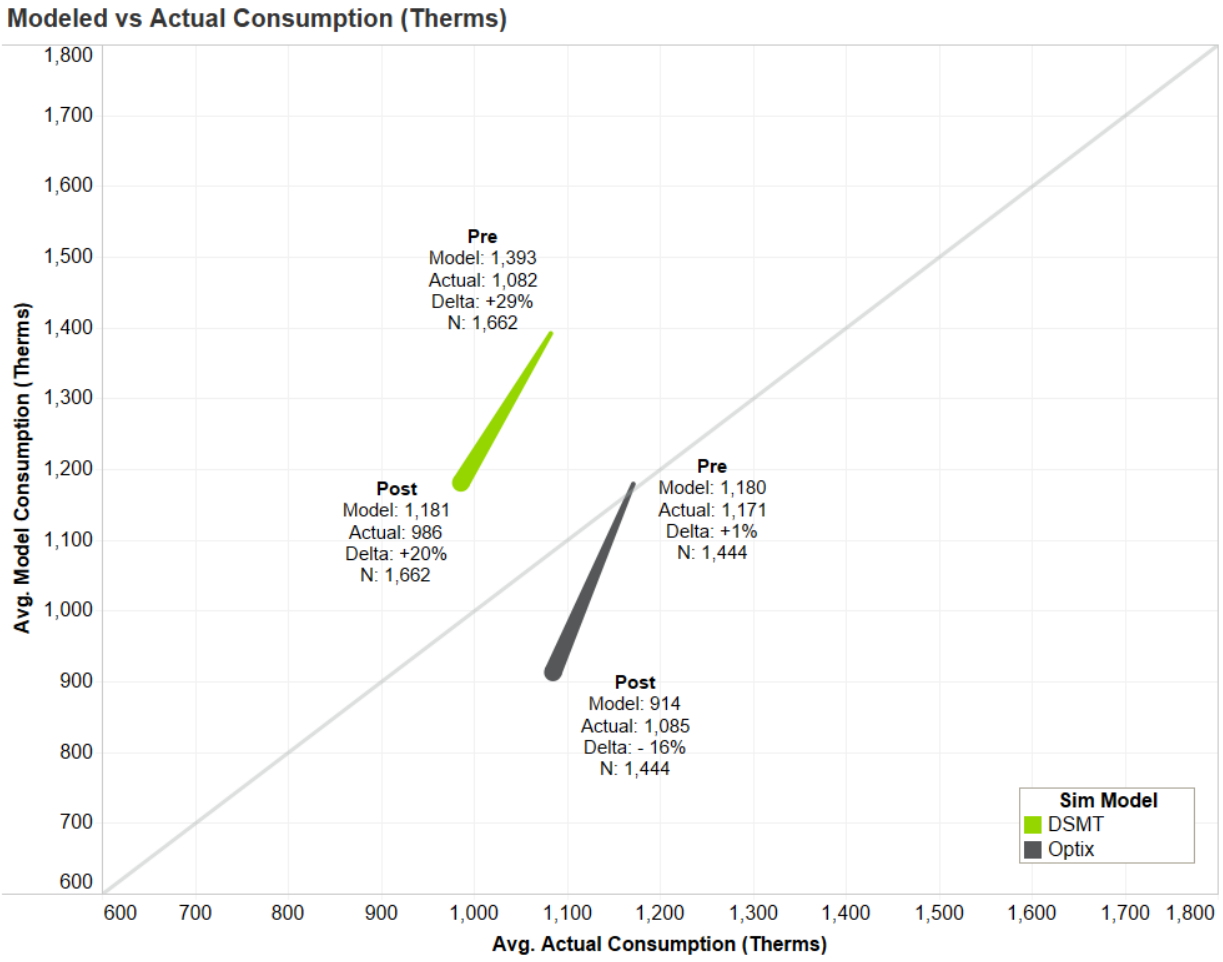
As shown below, CLEAResult's DSMT model, on average, overpredicted actual participant's annual energy consumption in the pre- and post-period by 29% and 20%, respectively. The consequence of overpredicting consumption is that savings, the difference in consumption between pre- and post-installation, is also overstated. For example, even if the DSMT accurately predicted that weatherization would reduce energy consumption by 10%, the absolute savings that PA's claim in their programs are inflated.

Looking at the average values for HES participants modeled using Optix tells a different story. Figure 3 confirms the model's accuracy in the pre-period (only 1% average difference from actual consumption), Optix underestimates consumption (again, relative to actual consumption) in the post-period by 16%. While the estimated consumption in the pre-installation period is more accurate, the net result is that the Optix model overstates savings because it underpredicts consumption in the post-installation period.

In addition to analyzing the modeled and actual consumption averages, our team reviewed the coefficient of variation[10] of those estimates. The team noted that Optix consumption estimates in our sample have a lower coefficient of variation (38% pre-, 37% post-) than the DSMT estimates (44% pre-, 41% post-), indicating that the Optix model results are marginally more precise. Similarly, the post- estimates have a lower coefficient of variation, indicating that they are marginally more precise that the pre- estimates.

[10] Coefficient of variation is defined as the ratio of standard deviation to average and provides a way to measure variability on a standardized scale.

Figure 3. Average Pre- and Post-Consumption Estimates by Software



3.2 Differences by Participant Type

Relevant Research Question: Is the disconnect between ex ante and ex post savings more pronounced for certain subsets of HES participants (e.g., single family detached vs. 2-4 units or participants that received different types of insulation)?

Key Finding: Both whole-home simulations less accurately estimated the consumption of older homes, which typically receive more insulation than newer homes.

The team also used the Comparative Analysis to assess whether the initiative-wide trends shown in the previous figures held when assessing specific subsets of HES participants. In other words, do the LV simulation models more accurately estimate consumption and predict savings for certain participants and/or building types? Conversely, are there particularly problematic participant and/or building types that are dragging down HES' overall realization rate.

Specifically, our team assessed four specific participant subsets:

1. Home Type (e.g., Ranch, Cape Cod, multi-unit)
2. Home Age
3. Home Size (i.e., square feet)
4. Weatherization Type

Again, because this assessment relied on the Comparative Analysis, the team's assessment focused on Eversource and National Grid participants. Figure 4 and Figure 5 assess differences in modeled versus actual consumption across all four participant/home types for both Eversource (CLEAResult's DSMT) and National Grid (EnergySavvy's Optix) participants.

Consistent with Figure 1, the Eversource results show a general overestimation across all four participant /home types. DSMT appears to overestimate consumption, in both the pre- and post-period, relatively consistently across the assessed home types and home sizes. However, the magnitude of the overestimation differs by home age. The model appears to estimate consumption less accurately for older homes; for example, DSMT modeled pre-program consumption for homes built between 1900-1920 is nearly 50% higher than actual consumption for those homes.

That said, building simulation model inaccuracy with older homes is not uncommon. [11,12,13] Older homes, by virtue of their age, are known to be more difficult to accurately model. The complications are multiple and often conflated: originally constructed before modern building codes, older fenestration, renovations to the original building envelope, and multiple cycles of mechanical equipment turnover. In short, older buildings are more complicated.

MODEL OVERPREDICTION: NOT UNCOMMON

Numerous other studies have shown that overprediction of energy use and savings by residential energy analysis methods is a common problem. Our team's literature review found the issue is particularly acute for older, poorly insulated homes containing aging mechanical systems. Of course, these are also the very homes where the most HES opportunities exist. Specifically, an NREL study found that simulation software overpredicted natural gas space heating by an average of 41% for homes built before 1960 compared to a 13% overprediction for homes built after 1989. This finding is largely consistent with this study's findings.

Oak Ridge National Laboratory also conducted a study comparing modeled space heating energy use with actual space heating energy (i.e., billing data). That study found overprediction was more common for the pre-period, leading to an overprediction of savings. However, the study also found post-retrofit modeled consumption within an (unspecified) acceptable range.

While, simulation models (per our team's literature review) are known to overpredict savings, it is important to reiterate that the models are important program delivery tools. The models allow assessors to consistently estimate energy consumption and savings that the assessors can relay to the participant before leaving their home. The models may be imperfect but investigating where they consistently deviate from actual consumption can improve their accuracy while maintaining their usefulness.

To this end, Figure 4 shows large disparities between actual and DSMT-modeled consumption for homes that received multiple measure and for older homes (pre-1940 in particular). Though the magnitudes are different for the Optix model, Figure 5 shows a similar theme in that homes with multiple measure

[11] Field Assessment of Energy Audit Tools for Retrofit Programs Edwards, D. Bohac, C. Nelson, and I. Smith NorthernSTAR Building America Partnership. https://bbe.umn.edu/sites/bbe.umn.edu/files/Field%20Assessment%20of%20Energy%20Audit%20Tools%20for%20Retrofit%20Programs.pdf

[12] Assessing and Improving the Accuracy of Energy Analysis for Residential Buildings B. Polly, N. Kruis, and D. Roberts. https://www.nrel.gov/docs/fy11osti/50865.pdf

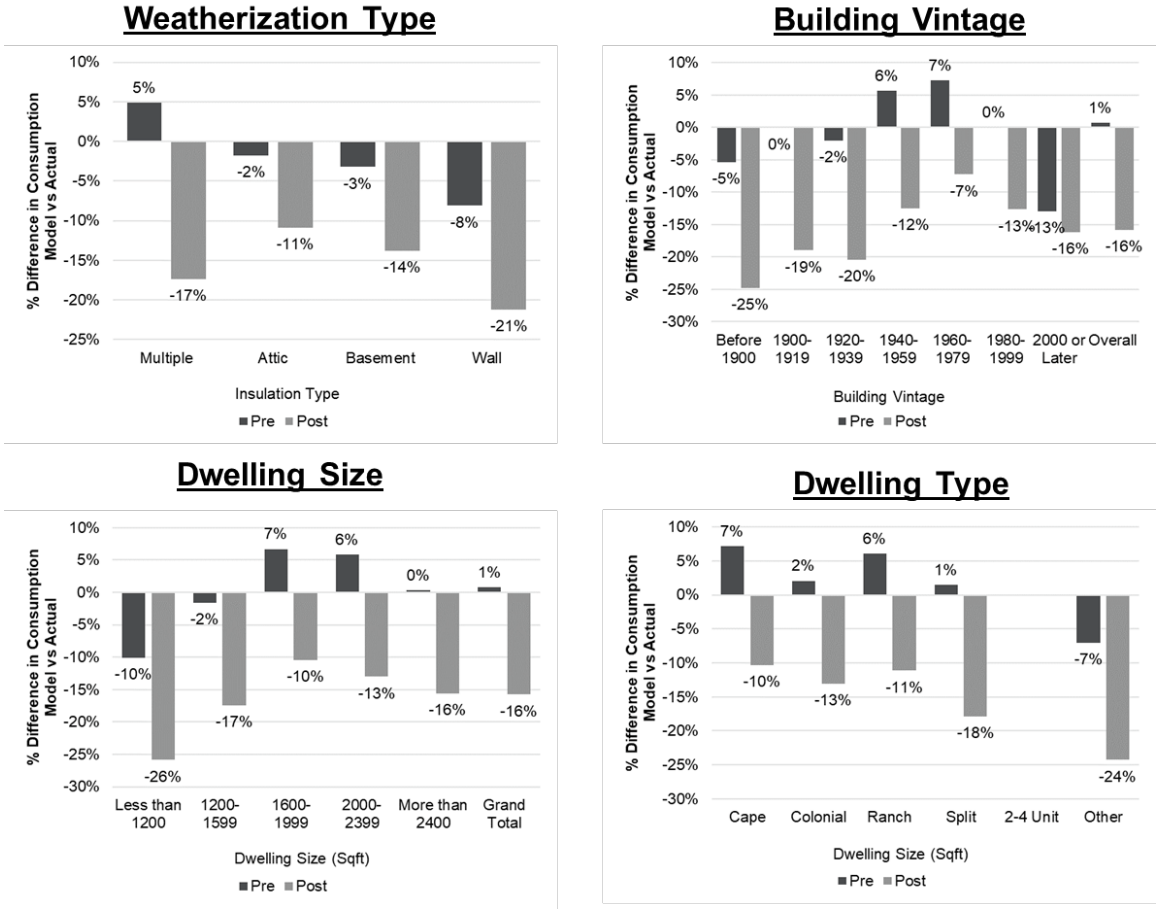
[13] Ternes, M.P. (2007). Validation of the Manufactured Home Energy Audit (MHEA). ORNL/CON-501 and the second part of the study referenced in Ternes, M.P; Gettings, M.B. (2008). Analyses to Verify and Improve the Accuracy of the Manufactured Home Energy Audit. ORNL/CON-506.

installations and older homes tend to have the largest discrepancies between actual and modeled consumption.

The team was not able to reliably quantify the influence of building age and multiple measures on modeled consumption with a statistical model because the two factors are correlated, however, the empirical data suggest two potential issues with the building simulations. First, this could indicate the models may not be accounting for the interactive effects between measures and as a result, overstate savings. Another plausible explanation is that the homes that received multiple types of insulation are also older homes and that in general, the models overpredict energy consumption in older homes due to inaccurate assumptions about shell quality (i.e., understanding the R-value of insulated area in the pre-period and/or overstating the effectiveness of added insulation in the post-period).

Figure 4. Comparison of Pre- and Post-Modeled and Actual Natural Gas Consumption: For Subsets of 2017 Eversource Weatherization Participants (n=1,662)

Figure 5. Comparison of Pre- and Post-Modeled and Actual Natural Gas Consumption: For Subsets of 2017 National Grid Weatherization Participants (n=1,444)



Note: There are no Optix-modeled dwelling units flagged as multifamily – any multifamily dwelling units are rolled into the "other" category

3.3 Model and Process Integrity

Relevant Research Questions: Are the tools and processes that LVs use robust? Do the LVs models rely on the best available? To what extent, if any, does the experience and approach used by the Energy Specialist (ES) or Home Performance Contractor (HPC) impact modeling accuracy?

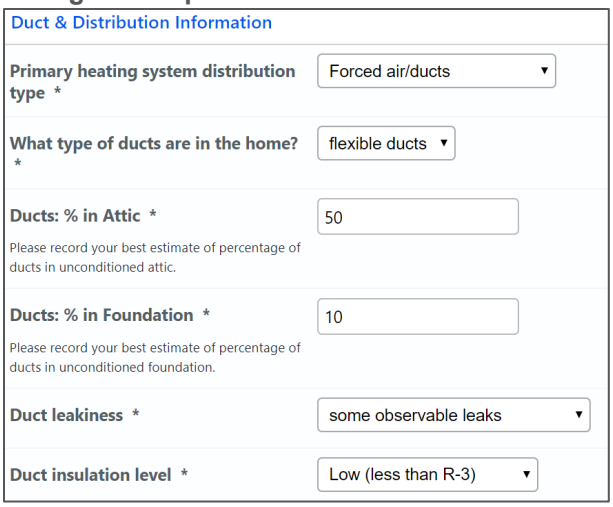
Key Findings: All three LVs use robust models that rely on MA-specific data. There does not appear to be meaningful differences in how ESs or HPCs use these tools in the field; most shared a similar data collection and modeling approach and reported they infrequently change model default values.

Based on our team's independent review of all four LV models, as well as our in-depth discussions with each LV's modeling team, we have determined that all four LV models are robust tools.
Each model effectively allows home energy assessors to quickly model a customer's home in real time. While not perfect, each model accounts for the critical factors that impact weatherization savings such as heating system efficiency, existing/proposed R-values, infiltration rates, square footage of insulation by

type, and area of the home. The team found that these models employ industry standard practices, and all contain the necessary functionality to estimate energy savings.

In some instances, the models go further. For example, our team assessed how the models accounted for a participant duct and distribution system. (It is possible the omission, or mischaracterization, of information regarding duct and distribution systems could impact weatherization savings estimates. As shown in Figure 6, EnergySavvy's Optix tool includes functionality to characterize a participant's duct and distribution system.

Figure 6. Optix Duct & Distribution Module



Our team determined that DSMT, the other whole-home simulation tool used in the state, includes similar functionality. Neither algorithmic approach directly addresses the condition of the participant's duct and distribution system.
Whether the models accounted for conditions of duct and distribution system is one example of the numerous modeling elements that our team explored as part of the model review, discussions with LVs, and interviews with Energy Specialists and HPCs. Table 2 summarizes the questions our team sought to answer for each model and, in doing so, documents each model's capabilities, default inputs, and visible assumptions. (As noted previously, we did not have access to background assumptions and engineering calculations behind the user interface for the models.).

Table 2: Summary of Savings Characteristics by Model

	RISE (Eplus)	CET	EnergySavvy	CLEAResult (EM Home/DMST)
PAs served	Columbia Gas, Cape Light Compact, Liberty Gas, and Unitil	National Grid	Eversource	Berkshire Gas
Are there any differences in the way the model is deployed across PAs?	No	N/A	N/A	N/A
Does the model use whole-home simulation or an engineering equation estimate Wx savings?	**Equation.** Uses algorithms to estimate savings for individual Wx measures		**Simulation.** Calculates pre- and post-program consumption for the home to estimate savings from Wx measures	
Is the model capable of capturing customer-specific energy consumption data?	**Yes.** Includes a free response field to add customer consumption data. However, this field does not currently factor into Wx calculations	**No**	**No**	**Yes.** Includes a free response field to add customer consumption data. However, this field does not currently factor into Wx calculations
Is the model calibrated using customer-specific energy consumption data?	**No**			
Does the model account for interactive effects between recommended measures when estimating Wx savings? (Includes effects between insulation types, as well as interactions with other end-uses like heating system)	**No.** Neither interactive effect type		**Yes.** Both interactive types	
Does the LV re-estimate air sealing and/or insulation-type specific savings when a customer installs a subset of recommended measures?	N/A		**Yes.** The LV re-estimates savings if a participant does not install all recommended measures	
Does the model factor in efficiency degradation for older heating systems?	**Yes.** Assessors are trained to estimate the in-situ efficiency of older heating systems or take measurements during an assessment			
Does the model account for distribution and duct losses	**No***		**Yes**	
Are final/reported air sealing savings calculated using blower door test data?	**Yes.** Contractors conduct pre and post installation blower door tests			

What weather data does the model use?	**PA-specific**, not customer-specific	Closest **NOAA weather station** to participant's address
Is the model capable of capturing thermostat setpoint variations by customer?	**No**	**Yes**, but default value provided and largely retained
Is the model capable of accounting for various building types/configurations (e.g., single family detached, side-by-side duplex, triple decker) when calculating Wx savings?	**No***	**Yes.** Assessors create building characteristics in the model specific to each home
Does the model account for secondary heating systems, such as additional furnaces/boilers or wood stoves when calculating Wx savings?	**No**	**Yes**

*As noted earlier in the report, both algorithm approaches include a "correction factor" in their Wx savings equations. Although this factor does not explicitly account for customer specific distribution systems, interactive factors, or home types, it does attempt to adjust modeled to account for unspecified factors.

There is significant variation between the four LV models, but in general, all four models are adequate tools to estimate weatherization savings during a home energy assessment. The following section details the team's key findings related to the integrity of the model and modeling process, as well as possible areas of improvement.

- **None of the models rely on participant-specific billing data.** Two of the LV models (DSMT and Optix) have the potential to add functionality to integrate customer-specific energy consumption data (including for delivered fuels), but CLEAResult and the interviewed Energy Specialists and HPCs reported that they do not currently take advantage of any such functionality. Currently, both whole-home simulation models rely on assumed total household annual energy consumption to estimate weatherization savings. In theory, it would be possible to replace this assumption with a specific participant's actual pre-program consumption. EPLUS and CET's models do not have the capability for this functionality, as they are algorithm based, and therefore calculate standalone savings for any proposed weatherization measure.) The lack of participant-specific data means that assessors present HES participants with estimated weatherization savings disconnected from their personal pre-program energy use. Interviewed Energy Specialists and HPCs shared with our team that customers often provide them with copy of their bill (an estimated three-quarters of the time) so the information is nearly always available. Other than a small number of assessors that mentioned irregularly using it to do a "sanity check" on the model outputs, this information commonly goes unused (i.e., the assessor does not modify the model or its outputs).

- **Assessors use models consistently.** Assessors generally populate and operate the four LV models consistently. Our team heard very little variation between the individual Energy Specialists and HPCs, as well as between LV models. For examples, interviewed assessors all described a similar approach to estimating existing insulation R-values based on BPI training, rules of thumb, and de-rating factors based on insulation condition. Some variability exists for "hard-to-reach" areas and crawlspaces, particularly if the assessor is time-constrained or does not plan to recommend insulation in the specific area anyway. Assessors' approach to estimating heating system efficiency also consistently followed one of two methods: measuring system efficiency directly or using a consistent estimation process based on the equipment's rated efficiency, age, and condition.

- **LVs using whole-home simulation models update them after the participant decides the measures they will install.** CLEAResult and EnergySavvy both update measure-specific savings estimates after the participant decides which recommended measure (or measures) they will install through HES. This post-decision model update is important as it ensures the LVs' modeled savings reflect the final set of installed measures and, therefore, appropriately account for interactive effects between different types of insulation (i.e., wall, attic and floor) and heating system. Since the algorithmic approach employed by RISE and CET involves modeling savings for each recommended weatherization measure individually (thereby not accounting for interactive effects), neither LV undertakes a similar post-decision update.

- **Contractors conduct pre- and post-blower doors tests.**[14] If a participant decides to weatherize through HES, the program undertakes a pre- and post-weatherization blower door test and uses this information to model the participant's total consumption (building simulation models only) and savings estimate (all LV models). Our team did not assess the accuracy of the testing process or the reasonableness of participant's test values, but this "test-in/test-out" approach likely produces high quality information for the models. While blower door tests have limitations (particularly for individual units in a 2-4-unit building), they represent the most robust way to calculate savings based on the options available.

[14] For all PAs except National Grid, RISE Energy Specialists also perform a blower door test during the assessment itself.

- **Weatherization-specific calculations have limited inputs.** The non-whole-home models do not account for the full range of factors that can impact consumption and savings. The weatherization-specific calculation methods do not include interactive effects, occupancy, thermostat setpoints, building type, lighting changes, or secondary heating systems, which can have large impact on energy consumption. However, while a whole-home model captures more customer-specific inputs, the weatherization-specific calculations do not necessarily result in a less accurate savings estimate.

- **Models account for secondary heating most of the time.** Three of the models, CLEAResult, CET, and EnergySavvy, incorporate secondary heating systems into consumption estimates, but RISE does not. Since weatherization savings estimates rely on accurately characterizing all of a home's heating systems, accounting for all forms of heating is necessary to accurately estimate energy savings.

3.4 Model Customization

Relevant Research Question: How does each LV customize their savings estimates to reflect a specific participating customer?

Key Finding: Energy Specialists and HPCs do not typically utilize customer-specific billing data when modeling a participant's home and estimating weatherization savings. However, assessors typically have access to participant's usage information (i.e., the participant shares a copy of their bill). In these instances, some assessors will use the bill to do a "sanity check" on the model's estimated savings.

As noted in the previous finding, all the LVs modeling tools—whether a whole-home simulation or weatherization-specific algorithmic model—are capable of detailed participant-specific customization. During our interviews with sampled assessors, we asked them the extent to which they adjusted various model parameters for each participant.

Most assessors shared that they always update a select set of priority model inputs, including home size, occupancy, insulation levels, and heating system type and efficiency. They correctly cited these inputs as critical for estimating the savings likely to results from weatherization. The assessors went on to say, largely due to limited time to perform the modeling while in customer's home, that they mostly use the model's default values for the rest of the potentially customizable fields.[15] For example, most assessors said that they did not input participant-specific heating and, when relevant, cooling temperature set points. Data provided by EnergySavvy supported this qualitative interview finding; it showed that nearly all participants (greater than 95%) were modeled using the Optix's default heating and cooling temperature set points (68° and 75°F, respectively). It is important to note that Optix's (and DSMT) default set points are reasonable, as well as consistent with temperatures used in the most recent evaluation to estimate weatherization savings for HES participants that heat with a delivered fuel using building simulation (i.e., when billing data was not available). [16]

[15] National Grid noted that they have recently worked with EnergySavvy to remove many default inputs from the Optix model so that assessors are required to input building- and customer-specific information.

[16] These default values are also similar to the thermostat set point identified by the broader residential evaluation team as part of the ongoing Residential Baseline Study.

3.5 Participant Behavior Change

Relevant Research Question: Are participants changing their behavior after being weatherized in a way that contributes to the disparity?

Key Finding: Rebound effect is not a meaningful contributor to the realization rate disparity observed in the previous two HES impact evaluations

Rebound effect (i.e., participants changing their behavior, such as increasing their heating temperature setpoint after weatherizing their home) is another potential contributor to the observed disparity between LV ex ante and evaluator ex post savings estimates. This type of post-participation behavioral change is also sometimes called "snapback".

Both terms refer to program savings that are eroded or completely lost when a participant changes their behavior in a way that leads to increased energy consumption following program participation. Behavior changes potentially impact LV realization rate because all four LV models assume that participants will behave similarly—for example, operate their heating system the same number of hours per year—before and after participating in HES. In fact, the only difference between the model's pre- and post-periods is the addition of the recommended insulation.

To assess the extent to which HES participants may be taking actions that produce a rebound or snapback effect, our team completed online interviews with 321 HES participants who weatherized their home in 2018.

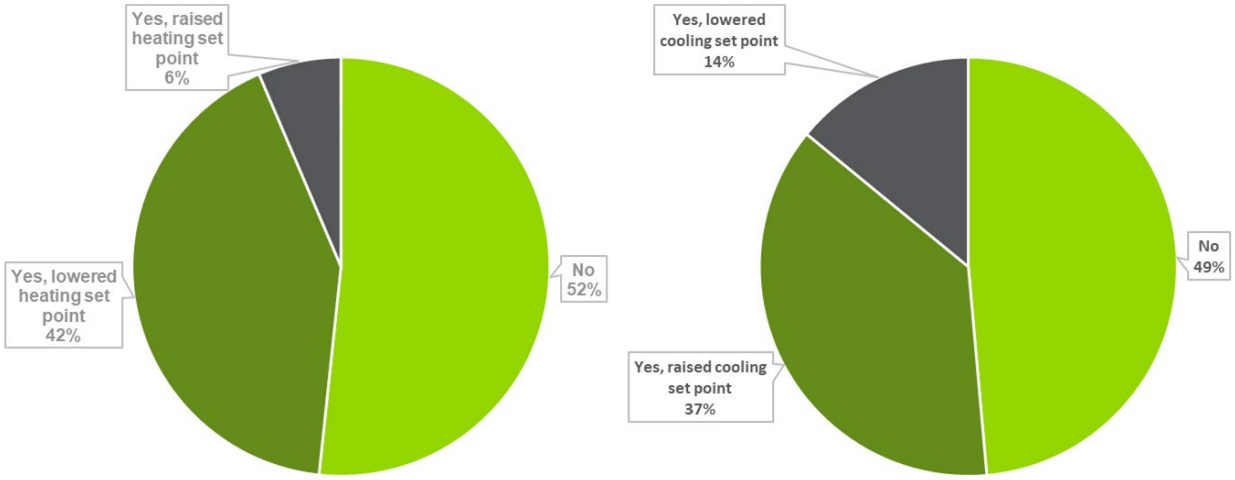
EXAMPLE OF REBOUND EFFECT

An example of the weatherization -specific rebound effect (or snapback) is a participant that decides to heat their home to a higher temperature (e.g., to 74° instead of 70° F) after having their home weatherized. The reason for such a behavior change is potentially directly related to their participation (e.g., the participant always thought their home was too cold but could not afford to heat it to a more comfortable temperature before to the program's efficiency improvements).

Our team's analysis of the survey results found little evidence of rebound effect that would contribute to the observed disparity between ex ante and ex post weatherization savings. As shown below, roughly half of surveyed participants (52% for heating and 49% for cooling) self-reported that they had not changed the heating or cooling set points following weatherization.

Of the participants indicating they had changed their set points after participating in HES, the majority noted a change that would result in lower consumption in the post-period and therefore greater savings. Specifically, 42% of participants indicating that they lowered their heating set point temperature (i.e., let their home stay cooler in the winter). When asked why they made this change, most participants selected the response: "*I don't need to set the heating temperature as high now because my home stays warmer because it's less drafty/insulated*." Similarly, more than a third (37%) of participants also reported that they raised their cooling temperature set point (i.e., allowed their homes to be warmer in the summer) after participating and provided a similar rationale.

Figure 7. Self-Report Changes to Heating and Cooling Temperature Set Points After Weatherization through HES



A small number of participants did report a behavior change that would lead to lower savings (relative to the assumption of consistent behavior in the pre- and post-period made by LVs); 6% for heating and 14% for cooling. However, these percentages are much smaller than the percentage of participants indicating a behavioral change that would result in greater savings

Our team also used the survey to ask participants if they changed the way they used their thermostat to heat or cool their home since participating in the HES weatherization program. The results were like above—about half said their behavior was unchanged, while those that did make a change nearly all indicated a greater reliance on their existing or new programmable and/or Wi-Fi thermostat, again indicating a behavioral change that would increase savings.

The team's analysis of the survey results did find one potential change in post-participation behavior that could potentially impact realization rates. Of the 321 surveyed participants, 124 (39%) indicated using a supplemental heating source (electric space heater, fireplace, wood stove, or oven) prior to weatherization. Changes in supplemental heating have the potential to impact realization rates because the billing analyses used by evaluators to estimate ex post savings are fuel specific. A billing analysis relying on a participant's natural gas energy consumption records would not, for example, account for any changes that participant made in the usage of an electrical space heater.

For this reason, our team asked these 124 participants whether they had made any changes in how they used their supplemental heat sources following weatherization. 40% of these participants indicated they had not made any changes in their usage of supplemental heating. A very small percentage (2%) reported increasing their usage. However, most (58%) reported decreasing their usage.

There are several possible reasons why a participant might decrease their use of a supplemental heating source. Since improvements to a participant's building envelope typically make home less drafty, weatherization can decrease the need for zonal, supplemental heat.

At first, it may seem that a decrease in the use of supplemental heating sources would be further evidence that rebound (or snapback) effect is not contributing to the HES realization rate disparity. However, the situation is not that simple: it is the rebalancing of primary and supplemental heating usage following weatherization that determines whether this behavioral change would lead to evaluators to over-report or under-report ex post savings.

For example, a tighter, less drafty home could lead a participant to decrease their use of supplemental electric heat and rely exclusively—and potentially more—on their primary fuel (natural gas for instance). This post-HES shift in heating source usage within the home would lead a natural gas-focused billing analysis to underestimate the total savings (across all fuels) in the participant's home. In such cases, the evaluator's natural gas weatherization realization rate would be artificially low relative to the total cross-fuel savings generated by the program's weatherization.

HOW BIG OF AN IMPACT ON REALIZATION RATES COULD SECONDARY HEAT SOURCES HAVE?

To explore the potential impact on realization rates, our team estimated the impact on evaluated savings for a participant that stops using their electric space heater **entirely** and replaces **all** the heating previously provided by their space heater with their primary, natural gas heating system.

In this example, our team assumed that a space heater uses 900 kWh per year (operating a 1.5 kW space heater 600 hours per year). Shifting all this previous electric heating usage to a participant's primary natural gas heating system would add approximately 25 therms to their post-HES natural gas annual consumption (accounting for differences in heating system efficiencies). Adding these 25 therms to the average weatherization savings (130 therms) found in RES 34 would increase the average savings to 155 therms.

However, this higher average savings would only be relevant for the subset of HES participants that used supplemental heat prior to HES (39% per the survey) **and** decreased their supplemental heating usage after weatherizing their home (58% per the survey). This subset represents less than a quarter of participants (22%, or 39% x 58%). Recalculating the savings using this higher average savings for 22% of participants (and using 130 therms for the rest) results in overall statewide average natural gas weatherization ex post saving of 136 therms.

Using the statewide ex ante savings from RES 34 (178 therms), hypothetically increasing statewide ex post savings to 136 therms would result in only a modest increase of the statewide realization rate (**from 0.73 to 0.76**).

4. RECOMMENDATIONS

Based on the key findings above, our team developed recommendations for the PAs as they look to answer the study's second core question: **What changes would improve vendor savings estimates?**

RECOMMENDATION 1. INTEGRATE PARTICIPANT CONSUMPTION DATA

Our model reviews and LV interviews determined that none of the HES LVs currently use a participant's pre-program energy consumption data to model their potential weatherization savings in a formal or systematic way. Not using participant-specific consumption data means the weatherization savings modeled for each participant are disconnected from that participant's actual energy usage. Our team believes this fundamental disconnect is likely a significant contributor to observed disparity in historical weatherization realization rates.

Since improving realization rates requires the HES LV models to accurately characterize energy consumption pre- and post-intervention, we recommend that PAs and LVs work together to formally integrate participant's actual historical energy consumption into the weatherization modeling process. We recognize that PAs and LVs have different legacy data management infrastructures and policies and, as such, also recognize the importance of providing each PA and LV with room for operationalizing this consideration.

Ideally, the LVs can load participant's consumption data into their models in an automated way prior to each assessment. This would prevent assessors from relying on participants to share a copy of their bill or from manually retrieving consumption data. However, if automated integration of consumption data into the modeling software is not technically possible—or requires significant time to implement—the PAs and LVs should take simpler, near-term steps toward integrating participant's consumption history. As an example, PAs could ensure that assessors have access to participant's consumption records and require them to review that information prior to the assessment.

RECOMMENDATION 2. FORMALIZE SANITY CHECKS

Our assessor interviews also determined that some assessors perform an informal "sanity check" to confirm that their modeled total consumption and savings values are reasonable using a copy of the participant's bill. However, this practice is assessor-specific and only occurs when a participant proactively provides a copy of their bill or can provide one upon request.

Consistent with the concept of better leveraging participant-specific consumption data to improve model accuracy, we recommend formalizing these sanity checks. There is variation in the capabilities of PA billing systems and LVs modeling systems, thus the sanity check could happen through either an automated or manual process. As an example, it may be possible to add an automated alert to the modeling software itself notifies the assessor if modeled consumption is beyond +/-10% of the home's actual consumption or the modeled weatherization savings exceed 35% of a participant's actual pre-program consumption. (Our team has offered placeholder values for these alert thresholds, but we suggest the PAs and LVs work together to select the appropriate values.) Alternatively, an example of a manual sanity check is a program requirement for the assessor to confirm that the model's consumption and/or savings fall within the designated range before the LV software produces a final assessment (and, if it does not, document

Whether implemented using an automated or manual approach, the formalization of these sanity checks will require all assessors to adopt this best practice and mitigate outlier savings estimates.

RECOMMENDATION 3: REVISIT EXISTING R-VALUE ASSUMPTIONS

Our team found meaningful differences between modeled and actual consumption for Eversource and National Grid participants across all types of insulation, particularly in older homes that received multiple types of insulation. (Our team was unable to complete a similar analysis for the other PAs.) Through the team's literature review and an interview with a building simulation industry expert, the team found anecdotal evidence that under-characterizing the R-value of the uninsulated building shell (especially in older homes) can be a source of overestimation of heating system energy use. Our team attempted to assess the default R-values associated with air films and the non-insulation portions of the building shell in each model (including wall assemblies, e.g. sheathing, siding, drywall), but we did not have access to this specific information. This and other background assumptions and/or engineering calculations were unavailable behind the model's user interface. For this reason, we recommend the LVs closely review all default R-value assumptions and consider using, at a minimum, a default of R-3 to reflect the thermal properties of a completely uninsulated wall.

APPENDIX A. COMPARATIVE ANALYSIS - ELECTRIC RESULTS

This appendix shows a similar analysis to that found in Section 3.1 and Section 3.2 of this document for participants with electric-heated homes.

The team's Comparative Analysis findings for electrically heated homes are similar to those in the body of this document for natural gas heated home, although the deviations are larger for electricity:

- In general, the simulation models overpredict electric consumption, and thus overstate the realized electric energy savings due to weatherization. The team found that modeled energy consumption was approximately twice that of actual consumption for HES weatherization participants, which is consistent with RES34 findings of a 54% realization rate for that measure.

- There are segments of homes – older homes and those that installed multiple weatherization measures – for which the DSMT and Optix simulation models perform worse relative to actual electric consumption.

The following information provides comparable figures to those provided in the body of the report, adjusted for electric energy consumption for participants with electric-heated homes.

Figure 8. Comparison of Pre- and Post-Modeled and Actual Electric Consumption: 2017 Eversource Weatherization Participants

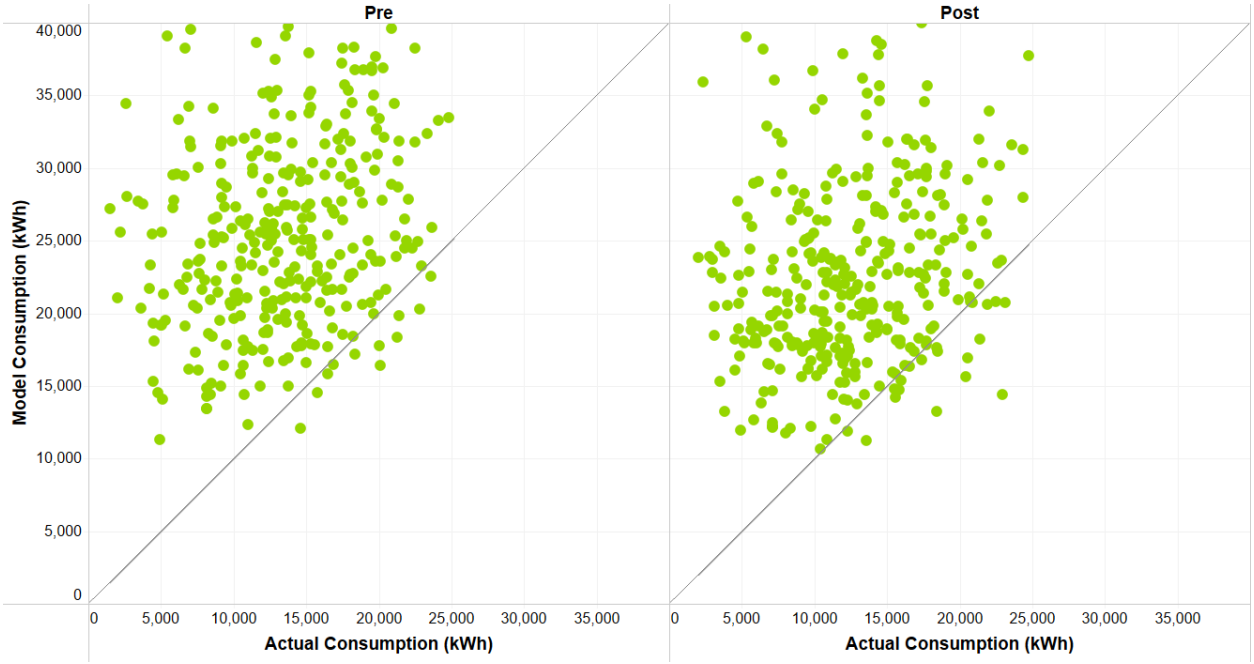


Figure 9. Comparison of Pre- and Post-Modeled and Actual Electric Consumption: 2017 National Grid Weatherization Participants

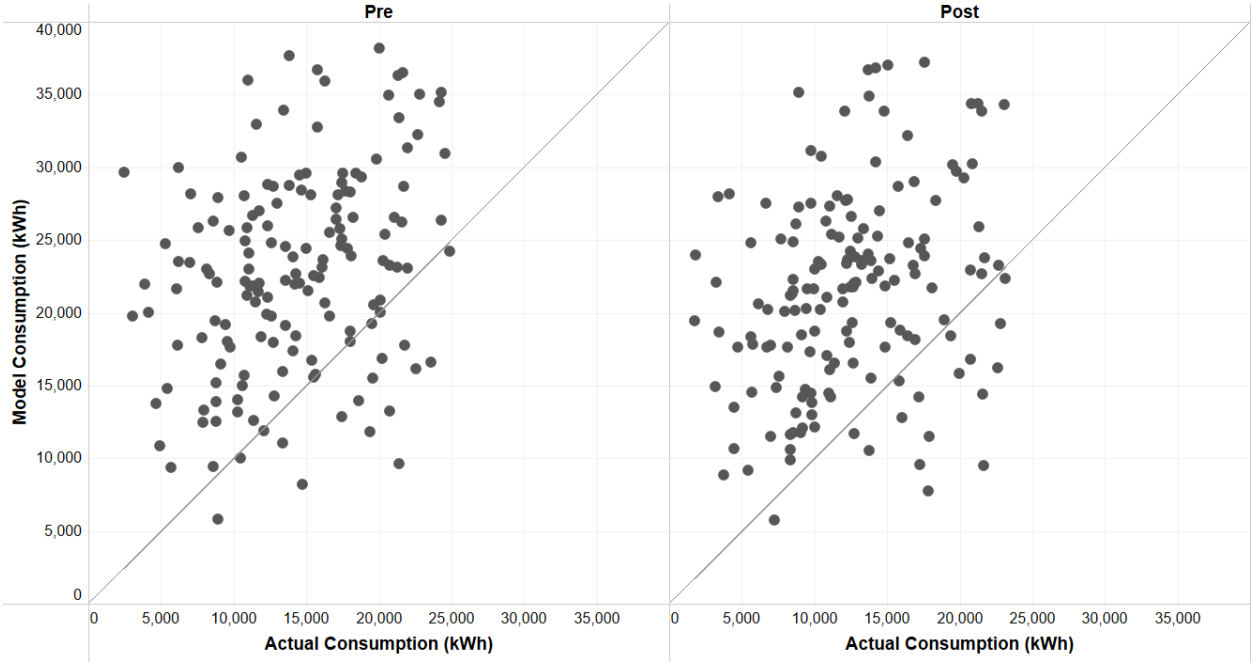


Figure 10. Average Pre- and Post-Consumption Estimates by Software

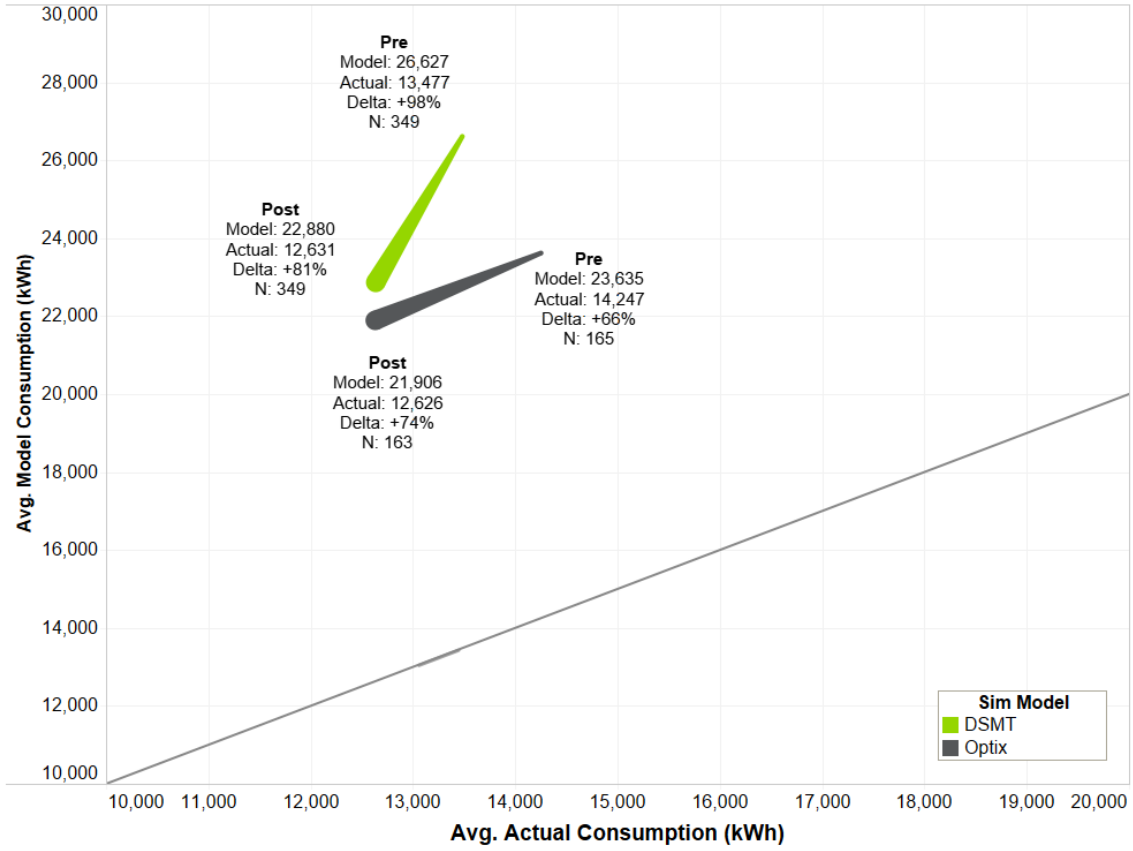


Figure 11. Comparison of Pre- and Post-Modeled and Actual Electric Consumption: For Subsets of 2017 Eversource Weatherization Participants (n=349)

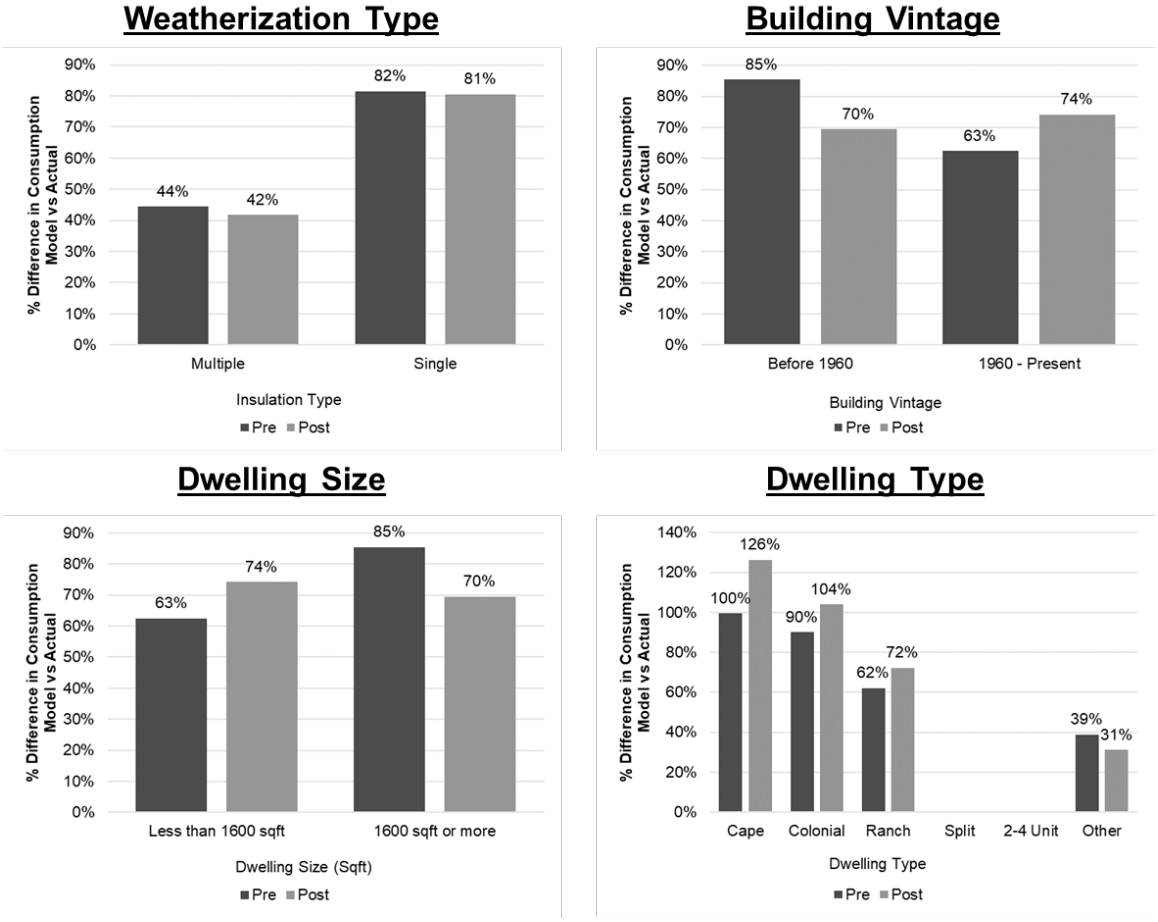



Figure 12. Comparison of Pre- and Post-Modeled and Actual Electric Consumption: For Subsets of 2017 National Grid Weatherization Participants (n=165)