Bonneville Power Administration

Impact Evaluation of FY2014-2015 Residential Insulation and Windows Measures

&

Analysis of FY2009-2011 Performance Tested Comfort System Air-Source Heat Pump Conversions and Performance Duct Sealing Data

February 2018

Developed by:

Jes Rivas, Pace Goodman, and Marilla Yaggie, Navigant Consulting, Inc.

Santiago Rodriguez-Anderson and Mike Baker, SBW Consulting

Developed for the Bonneville Power Administration

Please refer questions to:
Allegra Hodges, ajhodges@bpa.gov, 503.203.3904
Phillip Kelsven, pakelsven@bpa.gov, 503.230.4396

NAVIGANT

Table of Contents

Executive Summary

This report presents the findings from a billing analysis evaluation of Bonneville Power Administration's (BPA) residential envelope measures that contributed to the FY2015 Unit Energy Savings (UES) portfolio. The residential envelope measures include insulation and window measures in single family and manufactured homes. This report also includes the findings from a billing analysis of the Performance Tested Comfort System (PTCS) measures in Appendix E.

Background

BPA, along with its utility partners, acquires energy savings from a full portfolio of energy efficiency programs and measures. The majority of BPA's total reported savings comes from UES measures. UES measures utilize a constant deemed savings value for each measure application. In 2015, the evaluation team developed a plan to gain insight into BPA's energy efficiency programs through impact evaluation. The evaluation team systematically selected UES measures for evaluation based on their contribution to BPA's annual energy efficiency savings and their perceived importance amongst stakeholders. In that process, the evaluation team identified envelope measures for billing analysis evaluation, which would provide the appropriate balance of evaluation rigor and resources for these measures. Envelope measures include insulation and windows measures. Insulation measures include attic, wall, and floor insulation and window measures include both upgrades from single and double pane windows to U22 or U30.

Impact Evaluation Objectives

The team identified the following objectives for this billing analysis evaluation:

- Evaluate the energy savings achieved for consistency with the reported savings and UES as of December 2016.
- With the available data, assess the cost effectiveness of the evaluated savings using ProCost[1] and the updated 7th Plan inputs.
- When possible and applicable, use available data to begin to understand the drivers for unexpected evaluation results.

Methodology

The evaluation team conducted a billing analysis to estimate the impacts achieved across the sampled measures. The team compared these impacts to the reported UES

[1] ProCost is a model developed by the Northwest Power and Conservation Council and is used by the RTF to estimate the cost effectiveness of efficiency measures.

values (i.e. those in use during the time of measure delivery) as well as to the most recent UES to understand whether reported savings were achieved and whether BPA's more current savings estimate better aligns.[2]

To ensure the billing analysis methodology was robust, the evaluation team shared drafts of the proposed methodology, held meetings and incorporated input from regional experts and stakeholders, including RTF contract analyst staff, the RTF Statistical Methods subcommittee and BPA's Residential HVAC team. The team also addressed lessons learned from previous efforts in the region, and piloted the approach using the PTCS data that BPA collected in 2013. Finally, the evaluation team did additional targeted investigation to look for drivers of low or high savings compared to the UES. More details on the additional investigation is available in Appendix F.

Results

As summarized in Table ES-1, the impact evaluation found lower than expected savings for windows, but savings close to the December 2016 UES values for insulation. The evaluation team found a realization rate of 98% across all insulation types and 68% across all window retrofit types.

Table ES-1: Reported and Evaluated Savings by Measure

Measure	Reported Savings per Site* (kWh)	Savings per Site based on UES Dec. 2016** (kWh)	Evaluated Savings per Site (kWh)	Realization Rate over UES Dec. 2016 (%)
Insulation	2,780	1,331	1,303	98%
Windows	4,018	1,320	893	68%

*UES value referenced on BPA deemed measure list during time of measure delivery.

**UES value referenced on December 2016 from RTF measure workbooks.

Source: Navigant analysis

Using ProCost and the evaluated savings values, the evaluation team estimated cost-effectiveness for insulation and window retrofits. The evaluation team found that insulation measures are cost-effective, while window retrofits are not cost-effective at the evaluated savings.

[2] The most recent UES values here refers to those listed in December 2016. Differences between reported UES and UES as of December 2016 are driven by the Regional Technical Forum's (RTF) recalibration of the envelope UES. The recalibration was done in an effort to increase accuracy rather than an effort to adjust savings in response to changes to the implementation manual.

Table ES-2: Cost-Effectiveness for Residential Weatherization UES Measures

Measure Group	Present Value of Benefits per Home	Present Value of Costs per Home	Total Resource Benefit/Cost Ratio
Insulation	$3,450.22	$1,190.18	2.90
Window Retrofits	$2,253.06	$3,143.68	0.72

*Non-participant savings are not included in the cost-effectiveness analysis.
Source: ProCost Analysis using 7th Power plan inputs

Findings & Recommendations

The evaluation team presents the following findings and recommendations for insulation retrofits:

- **The evaluation team found that across all sites with insulation measures installed (attic, wall, and floor), the insulation measures are cost-effective using the evaluated savings.**
- **Navigant recommends that BPA continue to use the most up to date UES from the RTF.** The billing analysis evaluation found a realization rate of 98% using the most recent UES for insulation measures, indicating that the UES is close to the actual savings.

The evaluation team presents the following findings and recommendations for window retrofits:

- **The evaluation team found that window retrofits across all sites are not cost-effective with a B/C ratio less than 1.0 using the evaluated savings.**
- **Navigant recommends that BPA continues to use the most up to date UES from the RTF.** The UES from December 2016 is closer to the evaluated savings than the reported UES.
- **Navigant recommends that BPA coordinate with the RTF about including these results when the RTF conducts its scheduled review of window UES.** The results of this analysis indicate that the current UES may be high.
- **Navigant recommends that BPA consider additional research to better understand the window measure savings**. Additional research could include secondary literature review or process evaluation, which could lead to a better understanding of how end-use customers are interacting with their window retrofits and ultimately some identification of opportunities for additional savings from window retrofits.

1. Introduction

This report provides the results from the impact evaluation of FY2014-2015 residential insulation and windows retrofit measures as outlined in the Bonneville Power Administration Unit Energy Savings (UES) Portfolio Evaluation Plan for CY2016 Activities. [3]

2. Background

Prior to launching this billing analysis, the evaluation team analyzed BPA's UES portfolio, identified measures for evaluation, and reviewed previously collected data to finalize an evaluation approach. In this section, the evaluation team discusses the background for this work.

2.1. FY2015 UES Portfolio Summary

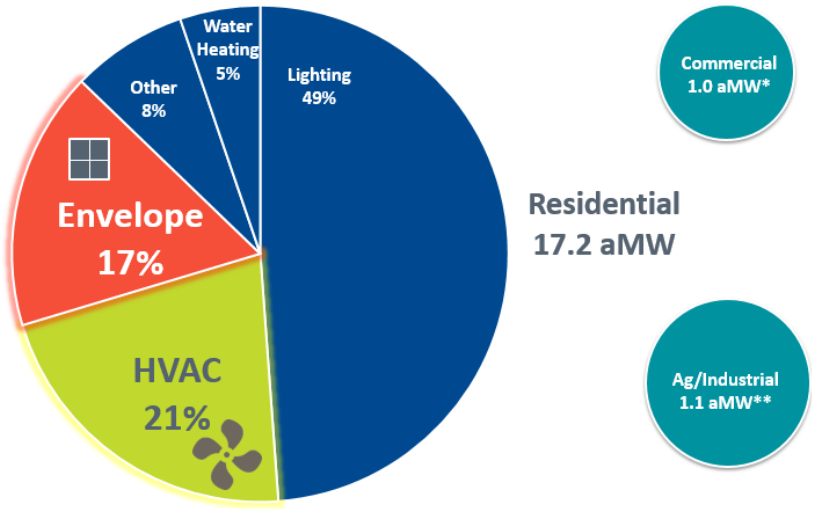
BPA, with its public power utility partners, acquires savings from a portfolio of energy efficiency programs and measures. The majority of BPA's total reported savings comes from UES measures,[4] which utilize a constant deemed savings value for each measure application.

UES measures fall into several categories of residential, commercial, and industrial equipment. As seen in Figure 1, HVAC and envelope measures are the second and third largest contributors to residential UES savings, providing the FY2015 UES residential portfolio with 6.6 MW.

[3] Navigant Consulting, Inc. April 2016. Bonneville Power Administration UES Portfolio Evaluation Plan CY2016 Activities. https://www.bpa.gov/EE/Utility/research-archive/Documents/Evaluation/BPA_UES_Evaluation_Plan_FINAL_04012016_V3.pdf

[4] In FY2016, 87% of the total savings in the BPA tracking database (Interim Solution 2.0 or IS2.0) were from UES measures.

Figure 1: FY2015 UES Portfolio Summary



* Savings from Energy Smart Grocers deemed measures are not included in this summary.
** Ag/Industrial value does not include savings achieved through the Scientific Irrigation Scheduling measure.
Source: Summarized from BPA's IS2.0 database, accessed 3/18/2016

2.2. BPA UES Portfolio Evaluation Plan for CY2016 Activities

In 2015, the evaluation team developed a plan to gain insight into BPA's energy efficiency programs through impact evaluation. The evaluation team systematically selected UES measures for evaluation based on their contribution to BPA's annual energy efficiency savings and their perceived importance amongst stakeholders. The evaluation team also identified the appropriate methods to evaluate these measures, while balancing strategic considerations including a measure's status, contribution to savings, uncertainty in claimed savings and programmatic importance.

As a part of the CY2016 evaluation planning, the evaluation team identified the following measures, amongst others, for billing analysis evaluation.

- **Envelope Measures:** residential insulation and window retrofits

The evaluation team determined that billing analysis would provide the appropriate balance of rigor and evaluation resources for these measures, given their importance to future program planning and contribution of savings for each measure (as seen in Figure 1).

2.3. FY2009-2011 PTCS Dataset

In 2013, BPA collected and analyzed billing data for Performance Tested Comfort System (PTCS) measures installed in FY2009 through FY2011. While the results of that analysis were not used directly, the data and lessons gathered are invaluable to the current evaluation effort in terms of 1) providing a rich data set to test methods, and 2) demonstrating that future billings analyses would benefit from (a) early and broad

engagement of regional stakeholders, (b) using a phased approach to pilot the analysis and results, and to create opportunities for additional research as warranted by findings; and (c) identifying the analysis results that will be used as evaluation findings before conducting the analysis, with parallel analysis to corroborate findings, to further develop and enhance analysis methodology.

Using this PTCS dataset, Navigant piloted the approach ultimately used for this evaluation, which is described in detail in

Billing Analysis Methods As a part of that work, the evaluation team developed results from the billing analysis of select 2009-2011 PTCS measure groups (i.e. performance duct sealing and ASHP conversions from electric forced air furnaces) and presents them in Performance Tested Comfort Systems Because of the age of the data, these results are not being treated and presented as an impact evaluation of today's program.

3. Methodology

This section provides an overview of the methodology used in the evaluation of BPA's residential insulation and window retrofit measures. The team provides additional detail in

.

3.1. Method Development

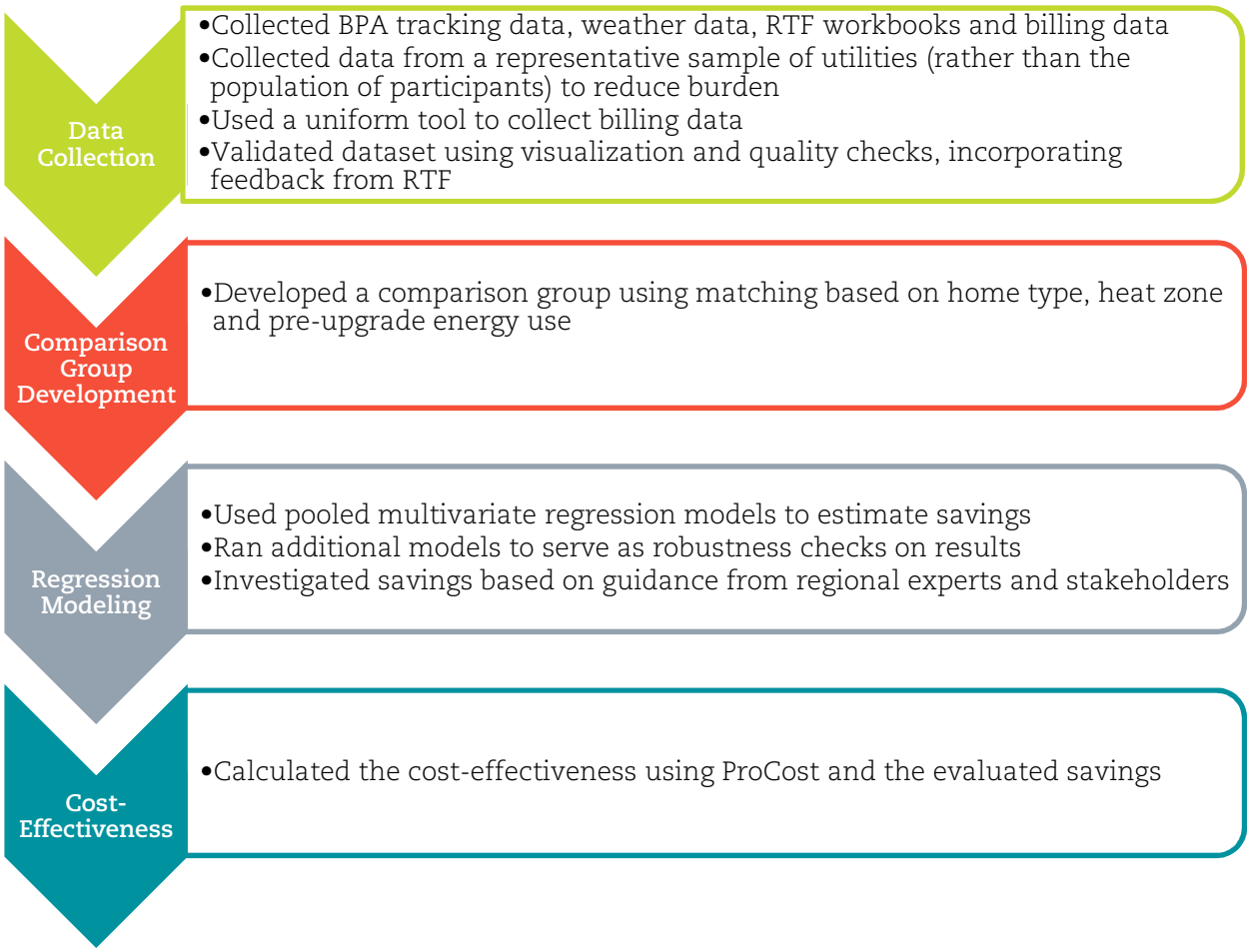
To ensure the methodology was robust, the evaluation team shared drafts of the proposed methodology, held meetings and incorporated input from regional experts and stakeholders (RTF contract analyst staff, RTF Statistical Methods subcommittee, BPA's Residential HVAC team). The team also addressed lessons learned from previous efforts in the region, and piloted the approach using the PTCS data that BPA collected in 2013.

3.2. Data Collection and Billing Analysis

Figure 2 outlines the overall process, and

provides the detailed methodology.

Figure 2: Outline of Billing Analysis Methodology



3.3. Additional Data Collection

As a final step, the evaluation team determined whether additional data should be included in the analysis to reduce uncertainty or to explain any unexpected findings. The evaluation team shared preliminary findings and a summary of available data with stakeholders to finalize this decision.

4. Results

In this section, the evaluation team provides the results for the residential insulation and window retrofit measures' achieved sample, evaluated energy savings and cost-effectiveness.

4.1. Sample & Representativeness

Table 1 details the sample design for this evaluation. Additional information can be found in Sections 5 and 6 of the CY2016 Evaluation Plan.[5]

Table 1: Draft 2016 Sample Size for the Residential Insulation and Window Retrofits

Measure Group	Strata	Assumed CV	Number of Utilities	Target Number of Projects*	Confidence and Precision Targets
Insulation	Large Contributors	0.8	4	Census of billing data for each utility**	90/15
	Medium Contributors	0.8	4		
	Small Contributors	0.8	2		
	Subtotal		**10**	**~1,700**	
Window Retrofits	Large Contributors	0.8	2	Census of billing data for each utility**	90/15
	Medium Contributors	0.8	10		
	Small Contributors	0.8	3		
	Subtotal		**15**	**~2,500**	

* This value represents the target number of projects for which the evaluation team requires usable data. In order to reach this number, the team will need to request billing data for roughly twice as many projects.

** Evaluation will target a census of energy consumption data for sampled utilities; a sample may be drawn where this is infeasible.

Source: Navigant Analysis

Table 2 includes the list of the 15 sampled utilities.

Table 2: Sampled Utilities for CY2016 Impact Evaluation of Residential Envelope Measures

Sampled Utilities
Lower Valley Energy, Inc.
Central Electric Cooperative, Inc.
Public Utility District No. 1 of Cowlitz County
Inland Power & Light Company
Clark Public Utilities
Northern Wasco County People's Utility District

[5] https://www.bpa.gov/EE/Utility/research-archive/Documents/Evaluation/BPA_UES_Evaluation_Plan_FINAL_04012016_V3.pdf

Sampled Utilities
Blachly-Lane Electric Cooperative
Public Utility District No. 1 of Grays Harbor County Washington
City of Cheney
Midstate Electric Cooperative, Inc.
Tacoma Power
Public Utility District No. 1 Of Snohomish County
Eugene Water & Electric Board
Lincoln Electric Cooperative, Inc.
Public Utility District No. 1 of Lewis County

Source: Navigant analysis

Table 3 lists all the sites with insulation measures included in the analysis by heat zone and home type. Almost 80% of the sites were single family homes in Heat Zone 1. While there were some manufactured homes, no multifamily buildings were included in the analysis.

Table 3: Distribution of Insulation Measures

		Single Family			Manufactured Homes		
	All Sites	HZ1	HZ2	HZ3	HZ1	HZ2	HZ3
Number of Sites	1,146	909	176	4	37	20	0

Source: Navigant analysis

Table 4 lists all the sites with window measures included in the analysis. Similar to the insulation measures, most of the sites were single family homes in Heat Zone 1. However, there were more manufactured homes included in the analysis for windows compared to the insulation measures. Again, multifamily buildings were not included in this analysis.

Table 4: Distribution of Windows Retrofit Measures

		Single Family			Manufactured Homes		
	All Sites	HZ1	HZ2	HZ3	HZ1	HZ2	HZ3
Number of Sites	1,378	970	328	0	61	19	0

Source: Navigant analysis

4.2. Savings

Overall, the impact evaluation found lower than expected savings for windows, but supported the December 2016 UES values for insulation (see Table 5). In the subsections below, the team provides additional context for the results of the impact evaluation.

Table 5: Reported and Evaluated Savings by Measure

Measure	Reported Savings per Site* (kWh)	Savings per Site based on UES Dec. 2016** (kWh)	Evaluated Savings per Site (kWh)	Realization Rate over UES Dec. 2016 (%)
Insulation	2,780	1,331	1,303	98%
Windows	4,018	1,320	893	68%

*UES value referenced on BPA deemed measure list during time of measure delivery.
**UES value referenced on December 2016 from RTF measure workbooks.
Source: Navigant analysis

4.2.1. Referenced UES

The evaluation team references reported savings (i.e. the UES values used during the time of measure delivery) and current savings (i.e. December 2016 UES values) for comparison to the evaluation findings. In each case, the UES reflect savings at the site as opposed to at the busbar. Savings at the busbar include the additional savings from avoided line losses, where savings at the site reflect savings from the energy consumed at site and after the customer meter.

Reported savings. Reported savings reference UES values included on the BPA deemed measure list during the time of measure delivery. These UES represent the best estimate of savings at the time and were used for BPA's reporting and tracking. The evaluation team provides these savings to facilitate a comparison of whether the reported savings were achieved.

Current savings. The evaluation team also compares evaluated savings to the UES included in the RTF measure workbooks as of December 2016. The evaluation team provides this comparison to help BPA understand how more current savings estimates align. The RTF recalibrated the envelope UES values, which drives the difference between UES values across this timeline. This difference reflects an effort to increase accuracy rather than an effort to adjust savings in response to changes to the implementation manual.

4.2.2. Insulation

While the sampled projects included a mix of attic, wall, and floor insulation measures, most of the insulation was installed in attics, followed by floors, and then walls (Table 6). All measures had a pre-existing conditions baseline.

Table 6: Insulation Measures Installed by Location

Measure	Number of Sites*
Attic Insulation	2,513
Floor Insulation	2,173
Wall Insulation	933
Total	**5,619**

*The site counts reflect the unfiltered data set, which includes sites that do not have adequate data for inclusion in the billing analysis data set.

Source: Navigant analysis

Table 7 lists the evaluated savings per site and shows the realization rate for the insulation measures for all sites and by home type and heat zone, when data permits. More detailed results are not presented where the sample size was small (i.e., less than 40) or where the error bounds on the savings estimate is greater than the estimate of average savings (i.e., precision greater than 100%). Single family home participants in heating zone 1 was the only subset of insulation participants that met these criteria.

In summary, for the insulation measures, the evaluation team believes the UES values in use by the RTF, at the time of this writing, are reflective of verified participant savings. However, the variation in savings (as indicated by the error bars in Table 7) is relatively high given the sample size is greater than 1,000.

Table 7: Insulation Measures Evaluated Savings and Realization Rate by Home Type and Heat Zone

Home Type/Heat Zone	Number of Participants	Savings per Site based on UES Dec. 2016 (kWh)	Evaluated Savings per Site (kWh)	Realization Rate over UES Dec. 2016 (%)	Precision
All Sites	1,146	1,331	1,303 ± 381	98 ± 29%	29%
Single Family/HZ1	909	1,313	1,145 ± 277	87 ± 21%	24%
Single Family/HZ2	176	Excluded due to precision greater than 100%			
Single Family/HZ3	4	Excluded due to less than 40 sites			
Mnf. Homes/HZ1	37	Excluded due to less than 40 sites			
Mnf. Homes/HZ2	20	Excluded due to less than 40 sites			
Mnf. Homes/HZ3	0	Excluded due to less than 40 sites (no sites)			

Source: Navigant analysis

4.2.3. Window Retrofits

All measures included in the analysis had a pre-existing conditions baseline. As shown in Table 8, the sample of projects included a mix of single pane and double pane baseline windows, with the majority being double pane. Additionally, the majority of window retrofits upgraded to U30 rather than to U22 windows.

Table 8: Site Count by Window Measure for Single Family Participants in Heat Zone 1*

Window Measure	Number of Sites
Single Pane to U22	12
Single Pane to U30	342
Double Pane to U22	11
Double Pane to U30	452

* Table 8 lists the number of sites by window measure type for single family homes in heat zone 1. Over 70% of the analysis data set is single family participants in heat zone 1 and these proportions are consistent with the population of participants.

Source: Navigant analysis

Table 9 lists the evaluated savings and realization rate for window retrofit measures for all sites and by home type and heat zone, when data permits. More detailed results are not presented where the sample size was small (i.e., less than 40 sites) or where the precision was unacceptable (i.e., greater than 100%). Single family home participants in heating zones 1 and 2 were the only subsets of window retrofit participants that met these criteria.

Table 9: Window Measures Evaluated Savings by Home Type and Heat Zone

Home Type/Heat Zone	Number of Participants	Savings per Site based on UES Dec. 2016 (kWh)	Evaluated Savings per Site (kWh)	Realization Rate over UES Dec. 2016 (%)	Precision
All Sites	1,378	1,320	893 ± 240	68 ± 18%	27%
Single Family/HZ1	970	1,340	882 ± 264	66 ± 20%	30%
Single Family/HZ2	328	1,319	1,090 ± 553	83 ± 42%	51%
Single Family/HZ3	0	Excluded due to less than 40 sites (no sites)			
Mnf. Homes/HZ1	61	Excluded due to precision greater than 100%			
Mnf. Homes/HZ2	19	Excluded due to less than 40 sites			
Mnf. Homes/HZ3	0	Excluded due to less than 40 sites (no sites)			

Source: Navigant analysis

The team also considered the baseline number of panes and the efficient-case U-value when evaluating savings. Table 10 lists the evaluated savings and realization rate for a subset of the window retrofit measures. All the measures included in this analysis are in single family homes in heat zone 1 or 2. In all cases, the evaluated savings are consistently lower than the deemed savings estimate based on UES as of December 2016.

Table 10: Window Measures Evaluated Savings by Baseline Number of Panes in Single Family Homes

Number of Panes/Heat Zone	Number of Participants	Savings per sqft based on UES Dec. 2016 (kWh/sqft)	Evaluated Savings per sqft (kWh/sqft)	Realization Rate over UES Dec. 2016 (%)	Precision
1 Pane to U30 (HZ1)	225	11	6 ± 3	54 ± 28%	52%
2 Pane to U30 (HZ1)	314	5	2 ± 2	46 ± 42%	92%
1 Pane to U30 (HZ2)	117	13	7 ± 7	52 ± 51%	97%

Source: Navigant analysis

In conclusion, the evaluation team finds the following,

- Savings are lower than expected
- Savings are consistently low or have poor precision across many available characteristics and data (e.g., home type, heat zone, baseline number of panes, efficient-case U-value)
- The error is relatively high given the sample size of over 1,000 participants

4.3. Cost-Effectiveness

Using ProCost and the evaluated savings values, the evaluation team estimated cost-effectiveness for insulation and window retrofits (see Table 11).

Table 11: Cost-Effectiveness for Residential Weatherization UES Measures

Measure Group	Present Value of Benefits per Home	Present Value of Costs per Home	Total Resource Benefit/Cost Ratio
Insulation	$3,450.22	$1,190.18	2.90
Window Retrofits	$2,253.06	$3,143.68	0.72

*Non-participant savings are not included in the cost-effectiveness analysis.
Source: ProCost Analysis using 7th Power plan inputs

The evaluation team found that insulation is cost-effective, but that window retrofits are not. Specifically, the combination of the low evaluated savings for windows

(RR=~68%) and their higher incremental cost drives their benefit/cost ratio below 1. Additional details are included in Appendix D:

5. Recommendations

Navigant provides program recommendations and future research recommendations in this section.

5.1. Program Recommendations

Navigant recommends the following:

- BPA continue to use the most current RTF UES values for residential insulation and window retrofits.
- BPA share the findings resulting from this impact evaluation of window retrofits with the RTF so that they may incorporate these findings into the next review of this measure.

5.2. Future Evaluation & Research Recommendations

Navigant recommends that BPA consider prioritizing future evaluation resources on measures other than insulation, because the findings here support the current insulation UES values. While there may be opportunities for program improvement or for insulation measures to achieve additional savings, the insulation measures are currently cost-effective and meeting the expected savings, while other measures are not.

Navigant recommends that BPA consider additional research for window measures, because the realization rate and benefit/cost ratio were lower than expected and because this measure is important to BPA's utility customers and stakeholders. To better understand the current UES values, BPA could consider conducting a secondary literature of window retrofit savings in other regions, document reviews and/or engineering reviews of the current UES. To better understand how end-use customers are interacting with this measure, BPA could consider surveying program participants. Lastly, to better understand how this measure is being delivered, BPA could consider conducting interviews with trade-allies.

Regarding future billing analysis research and based on the findings from this work, Navigant recommends BPA (1) incorporate stakeholder input on methodology before collecting data and before conducting analysis; (2) use a phased approach to pilot the analysis and results, and to create opportunities for additional research as warranted by findings; and (3) identify the analysis results that will be used as evaluation findings before conducting the analysis, with parallel analysis to corroborate findings.

Appendix A: Glossary

Coefficient of Variation (CV)

A normalized measure of dispersion of a probability distribution and defined as the ratio of the standard deviation, σ, to the mean, μ:

$$c_v = \frac{\sigma}{\mu}$$

Delivery Verification - RTF Guidelines stipulate that Impact Evaluation may be accomplished using delivery verification to estimate savings for Proven UES (Unit Energy Savings) measures, i.e., savings equal the verified delivery quantity multiplied by the proven UES savings value. Delivery verification may also be useful in measure development and providing feedback to programs. The RTF Guidelines provide the following additional definition:

> *"Delivery verification involves physical inspection of measures or documentation of measures at the location where the program operator delivers them. For measures delivered to an end use, this involves collecting data from the end user facility to confirm that equipment conforms to the measure specifications. For measures delivered upstream of the end use, for example efficient bulbs sold through retailers, this might involve inspection of retailer or end user records of bulb sales or purchases."*[6]

Evaluation Measure Group - To design an efficient evaluation, the evaluation team defined subsets within sectors as a group of measures that have similar end-uses, measure statuses and/or that use similar program delivery method.

Impact Evaluation

Impact evaluation is used to estimate savings from energy efficiency measures. According to the RTF Guidelines, "program impact evaluations estimate savings from a period of program operation. Program impact evaluations involve the analysis of a reliable sample of program participants (and possibly non-participants) to determine the savings." The RTF Guidelines generally refer to evaluation of a portfolio or program, but are flexible in how evaluators define "program."

Measure Status - In the RTF Guidelines, a measure's category defines the savings estimation that should be used to evaluate savings. The RTF approves four measure categories within the UES portfolio; Proven, Small Saver, Provisional and Other.

[6] Details of the delivery verification strategies included in the 2016 UES evaluation approaches are discussed in detail for each domain in the Appendices.

Other UES

This includes measures that fall into the RTF-Small Saver and Planning categories, as well as UES measures that have been created by program operators but are not recognized by the RTF, such as BPA-qualified measures. Savings estimation methods for these measures require conducting one or more studies that may require site-specific data collection and analyses.

Realization Rate

The term is used in several contexts in the development of reported program savings. The primary applications include the ratio of project tracking system savings data (e.g., initial estimates of project savings) to savings that (1) are adjusted for data errors and (2) incorporate evaluated or verified results of the tracked savings. In the Updated Guidelines, the realization rate does not include program attribution.

Relative Precision

Measures the expected error bound of an estimate on a normalized basis. It must be expressed for a specified confidence level. The relative precision (*rp*) of an estimate at 90% confidence is:

$$rp = 1.645 \frac{cv}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}$$

where *n* is the sample size, *N* is the population size, and the coefficient of variance is *cv* = standard deviation / estimate mean value. The square root expression at the end of the equation is the finite population correction factor, which becomes inconsequential and unnecessary for large populations.

RTF Proven

These are measures for which the RTF has determined that savings estimation methods are proven and reliable.

Savings Realization Rate (RR)

The ratio of the field of evaluation energy savings to the program's claimed savings. The RR represents the percentage of program-estimated savings that the impact evaluation team estimates as being achieved based on the results of the evaluation M&V analysis.

Savings Validation

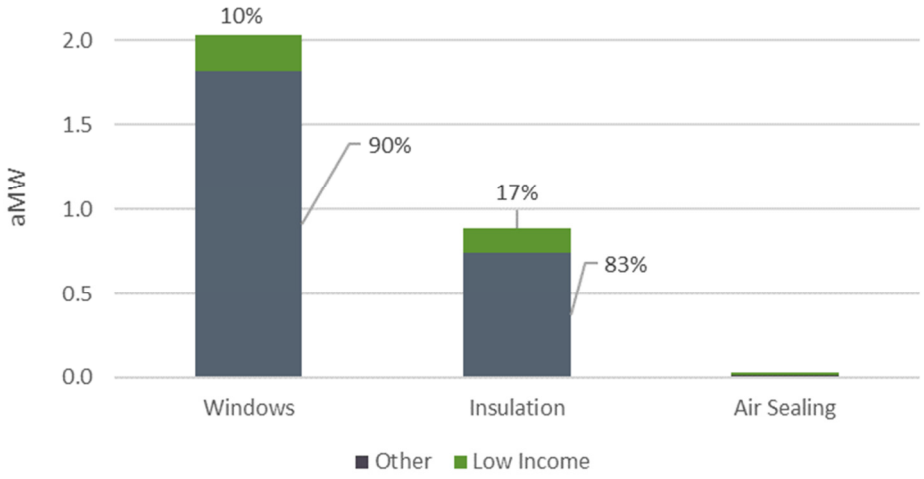
Savings validation uses impact evaluation to provide a comparison of savings for a measure or group of measures to the deemed UES values. For the purposes of this document, existing measure savings validation is considered a measure development activity, in that it informs savings estimates associated with a measure. If the savings

validation shows a significant deviation from the deemed savings estimates, additional measure development may be needed.

Appendix B: Measure Details

Figure B-1 shows the breakdown of TAP level energy savings for Residential Envelope Domain for Low Income measures and other (Non-Low Income) measures.

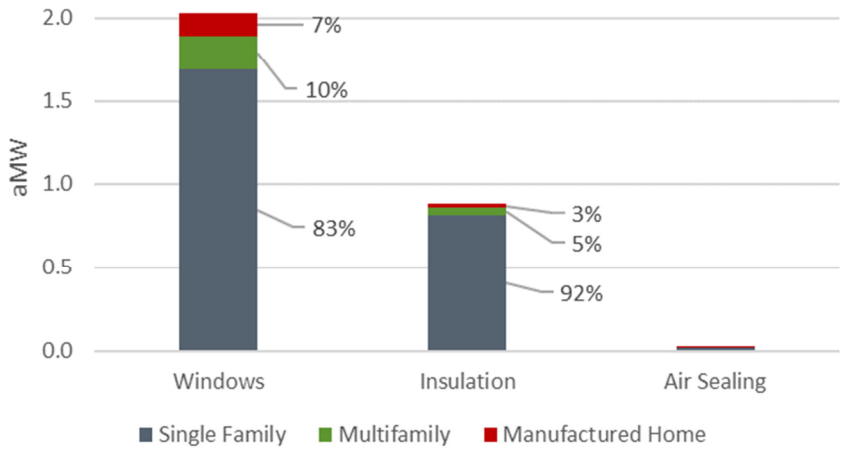
Figure B-1: Residential Envelope Domain Savings – Impacts of Low Income for each TAP (FY2015)



Source: Navigant analysis of measures reported into the BPA IS2.0, summarized from 3/18/2016 IS2.0 data pull

Figure B-2 shows the breakdown of TAP level energy savings for Residential Envelope Domain for different housing types.

Figure B-2: Residential Envelope Domain Savings – Savings by Residence Type for each TAP (FY2015)



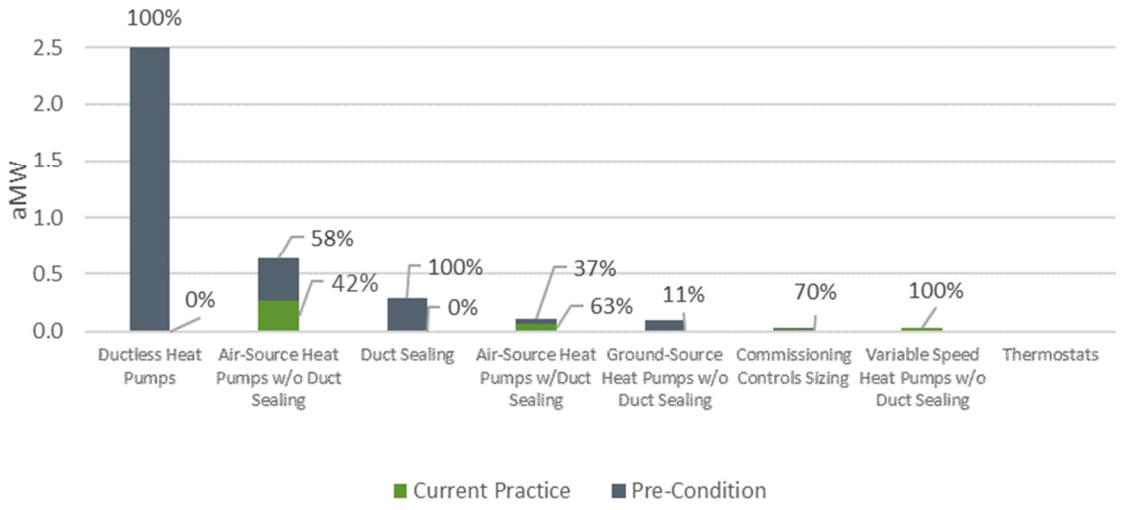
Source: Navigant analysis of measures reported into the BPA IS2.0, summarized from 3/18/2016 IS2.0 data pull

Figure B-3 show the breakdown of TAP level energy savings for Residential HVAC Domain for Current Practice baseline and Pre-Condition baseline.

A current practice baseline is characterized by current market practice or the minimum requirements of applicable codes or standards, whichever is more efficient. New construction and major renovations that are covered by codes and standards use this baseline.[7]

A pre-conditions baseline is used when the "measure-affected equipment or practice still has remaining useful life." In other words, the savings reflect improvement over the condition of the home before the measure was installed.

Figure B-3: Percent of TAP-Savings by Baseline (FY2015)



Source: Navigant analysis of measures reported into the BPA IS2.0, summarized from 3/18/2016 IS2.0 data pull

[7] *More information on the Current Practice baseline can be found here: https://www.bpa.gov/EE/Policy/IManual/Documents/July%20documents/9_BPA_MV_Absent_Baseline_Application_Guide_May2012_FINAL.pdf*

Appendix C: Unit Energy Savings (UES)

In this appendix section, the evaluation team discusses UES, the difference between reported UES and UES as of December 2016, and outlines our approach for assigning UES.

C.1 Unit Energy Savings (UES)

UES reflect the expected savings per unit from rebating certain efficiency measures. The RTF provides these values as deemed savings values for program administrators to plan for future years and to track their rebated savings. On pre-established timelines, the RTF reviews UES and may update values based on evaluation findings and other research.

For example, insulation is typically rebated per square foot installed, and as such, the RTF provides UES for insulation as annual energy savings per square foot of insulation installed.

C.1.1 Reported UES vs UES as of December 2016

When the RTF updates UES, the updated values only affect future years and the changes are not retroactive to previously rebated projects. As such, there is generally a delay from when the RTF updates UES to when program administrators use the updated UES as the value in program tracking and planning. In this evaluation, the UES were recently updated and the evaluation team was specifically interested in whether the updated UES were applicable to the BPA programs. As a result, the evaluation team put a greater emphasis in comparing the evaluated savings to the UES as of December 2016 than to the UES at the time the measures were installed.

C.1.2 Assigning UES

While reported UES were readily available in BPA's tracking data, Navigant had to assign UES as of December 2016 to each site in order to compare the evaluated savings to the most recent UES.

UES were assigned to each site included in the billing analysis by mapping the individual characteristics of each home and measure to the variables that defined a given UES value. Table C-1 identifies the relevant variables for each measure.

Table C-1: Variables used to assign UES by measure analyzed

Installation Site Variable	Insulation	Windows	Heat pump conversion	Heat pump conversion w/duct sealing	Performance Duct Sealing
Home type	X	X	X	X	X
Heating zone	X	X	X	X	X
Cooling zone			X	X	
Heating type (heat pump, electric resistance, electric forced air furnace, or generic "electric heat"- specified by utility)	X	X	X	X	X
Count of measures installed	X	X	X	X	X
With or without central air-conditioning			X	X	
Insulation Location (floor, attic, or wall)	X				
Existing insulation estimated R value determined by the use of an insulation identifier tool	X		X	X	
Base and retrofit window frame type		X			
Base and retrofit number of window panes		X			

In some cases, the team did not have complete information on all the variables necessary to assign UES. In these cases, the team used an average of savings based on the information available. For example, if it wasn't clear whether the home had central air conditioning, the team would average the savings for the UES that aligned with the site for all other variables. Also, the team did not have any information on the insulation type for the ASHP conversions and as a result averaged the UES across the savings provided for "Good," "Fair" and "Poor" insulation.

Appendix D: Cost-Effectiveness

The evaluation team used the RTF model ProCost[8] to estimate the lifetime sum of costs and benefits for each sampled measure/measure group. This model implements the Total Resource Cost (TRC) methodology which accounts for "all the costs of a measure with all of its benefits, regardless of who pays those costs or who receives the benefits"[9]. ProCost[10] outputs the discounted sum of costs and benefits over a measure's life.[11]

The team used the approach described in this report to estimate savings and calculate realization rates comparing evaluated to reported savings for the sampled measures. Data not provided by the program was taken from corresponding measures in RTF measure workbooks. This data includes annual Non-Electric Benefits (NEBs) such as O&M costs, and gas benefits from implementing measures.

To calculate the Total Resource Cost test (benefit divided by costs) for each measure group and for the portfolio, the team used the sample case weights to calculate an appropriately weighted sum of costs and benefits. The team also calculated the Total Resource Cost test for each sampled measure excluding any non-electric benefits.

For each case, a matching RTF measure or closest similar RTF measure was selected from the corresponding measure workbook. Each of these measures inputs were reformatted from their existing formats to match the format of the 7th power plan workbooks. Measure costs were taken directly from the BPA measure list (Capital Costs Column), which were developed by the RTF. Evaluated measure savings were used as the ProCost measure savings input. When measures had more than one energy savings component, the realization rate between the Deemed savings value and the RTF savings value was applied uniformly to each component (Example A)

Example A:

$$Realization\ Rate = \frac{BPA\ Unit\ Measure\ Savings}{Evaluated\ Unit\ Measure\ Savings}$$

$$ProCost\ Measure\ Heating\ Savings = RTF\ Unit\ Measure\ Heating\ Savings * Realization\ Rate$$

$$ProCost\ Measure\ Cooling\ Savings = RTF\ Unit\ Measure\ Cooling\ Savings * Realization\ Rate$$

The team then ran ProCost to determine the 7th plan costs and benefits for each measure. Analysts then weighted cost-effectiveness for a given measure by the claimed and evaluated measure quantities (EQ 1, EQ2).

[8] ProCost is a model developed by the Northwest Power and Conservation Council and is used by the RTF to estimate the cost-effectiveness of efficiency measures.
[9] From the 6th Power Plan.
[10] ProCost uses a slightly different busbar factor than the one used by BPA, which is also the one we have used throughout this report to show reported and evaluation savings. The ProCost busbar factor is 1.09066 and the BPA busbar factor is 1.09056.
[11] The average busbar factor used in this ProCost model is 1.075. For FY2015. the busbar factor used for BPA's residential lighting Retail and By-Request measures varied between 0.917 and 1.167.

$EQ1$: $Total\ Measure\ PVcost = Unit\ PVcost * Claimed\ Measures$

$EQ2$: $Total\ Measure\ PVbenefit = Unit\ PVBenefit * Verified\ Measures$

Finally, the team calculated the cost-effectiveness for a single measure or measure group using the following equation (EQ3).

$EQ3$: $Benefit\ Cost\ Ratio\ = \frac{\sum Total\ Measure\ PVbenefit}{\sum Total\ Measure\ PVCost}$

Appendix E: Performance Tested Comfort Systems

In 2013, BPA collected and analyzed billing data for Performance Tested Comfort System (PTCS) measures installed in FY2009 through FY2011. While the results of that analysis were not used directly, the data and lessons gathered are invaluable to the current evaluation effort in terms of 1) providing a rich data set to test methods, and 2) demonstrating that future billings analyses would benefit from early and broader engagement of regional stakeholders to further develop and enhance analysis methodology.

Using this PTCS dataset, Navigant piloted the evaluation approach described in

Billing Analysis Methods and produced results for several PTCS measures. Although these measures were not included in evaluation planning, the evaluation team developed results from the billing analysis of performance duct sealing and ASHP conversions from electric forced air furnaces. Specifically, the evaluation team provides results for the measure applications for which the methodology applies, i.e. those with pre-existing baselines. The evaluation team worked with RTF staff to document the changes to these measures over time, which is documented in Appendix G. Because of the age of the data, these results are not being treated and presented as an impact evaluation of today's program.

Table E-1 provides the average reported savings per site, the average savings per site based on December 2016 UES values, and the evaluated savings per site. The RTF has worked with these results, and they are aware of these findings.

Table E-1: Reported and Evaluated Savings by Measure from 2013 Data

Measure Group	Reported Savings per Site (kWh)	Savings per Site based on UES Dec. 2016 (kWh)	Evaluated Savings per Site (kWh)
Performance Duct Sealing	888	986	322
ASHP Conversions	3,932	6,006	3,705
ASHP Conversions with Duct Sealing	6,285	5,591	3,106

Source: Navigant analysis

Navigant provides more detailed results in the subsections below.

E.1 Sample & Representativeness

The PTCS dataset includes customers from the 42 utilities listed in Table E-2.

Table E-2: Sampled utilities for PTCS dataset

Sampled Utilities		
Ashland	Flathead Electric Coop	Monmouth
Benton Co. PUD #1	Franklin Co. PUD #1	Nespelem Valley Electric
Benton REA	Grant Co. PUD #2	Northern Lights
Big Bend Electric	Grays Harbor Co. PUD #1	Northern Wasco PUD
Central Electric Coop., Inc.	Inland Power & Light	Peninsula Power & Light Inc.
Central Lincoln PUD	Klickitat Co. PUD #1	Port Angeles
Clallam Co. PUD #1	Kootenai Electric Coop	Ravalli Electric Coop
Clark Co. PUD #1	Lane Electric Coop., Inc.	Richland
Clearwater Power Co.	Lincoln Electric Coop	Skamania Co. PUD #1
Columbia River PUD	Mason Co. PUD #1	Springfield
Consumer's Power, Inc.	Mason Co. PUD #3	Tacoma Power
Cowlitz Co. PUD #1	McMinnville	Tillamook PUD
Douglas Electric Coop., Inc.	Midstate Electric Coop	
Emerald PUD	Mission Valley Power	
Eugene (EWEB)	Missoula Electric Coop	

Source: Navigant analysis

Table E-3 lists the number of sites by home type and heating zone that are included in the final results. Most sites are in Heating Zone 1. Additionally, performance duct sealing measures were installed in more manufactured homes than in single family homes and vice versa for ASHP conversions.

Table E-3: Count of Sites Included in Final Results

Home Type	Heat Zone	HP Conversions	HP Conversions w/ Duct Sealing	Performance Duct Sealing
Single Family	1	292	502	899
	2	107	135	123
	3	-	-	4
Manufactured Home	1	9	160	3,375
	2	6	32	1,524
	3	-	-	262

Source: Navigant analysis

E.2 Savings

The next two sections provide the savings results separately for ASHP conversions and performance duct sealing.

E.2.1 ASHP Conversions

The team found ASHP conversion savings were lower than expected. Table E-4 and Table E-5 list the evaluated savings by home type and heat zone for ASHP conversions with and without duct sealing respectively. More detailed results are not presented where the sample size was small (i.e., less than 40 sites) or where the precision was unacceptable (i.e., greater than 100%).

Table E-4: ASHP Conversion Measures Evaluated Savings by Home Type and Heat Zone

Home Type/Heat Zone	Number of Participants	Savings per Site based on UES Dec. 2016 (kWh)	Evaluated Savings per Site (kWh)	Realization Rate over UES Dec. 2016 (%)	Precision
All Sites	414	6,006	3,705 ± 587	62 ± 10%	16%
Single Family/HZ1	292	6,061	4,178 ± 622	69 ± 10%	15%
Single Family/HZ2	107	5,966	2,434 ± 1430	41 ± 24%	59%
Single Family/HZ3	0	Excluded due to less than 40 sites (no sites)			
Mnf. Homes/HZ1	9	Excluded due to less than 40 sites			
Mnf. Homes/HZ2	6	Excluded due to less than 40 sites			
Mnf. Homes/HZ3	0	Excluded due to less than 40 sites (no sites)			

Source: Navigant analysis

Table E-5: ASHP Conversion with Duct Sealing Measures Evaluated Savings by Home Type and Heat Zone

Home Type/Heat Zone	Number of Participants	Savings per Site based on UES Dec. 2016 (kWh)	Evaluated Savings per Site (kWh)	Realization Rate over UES Dec. 2016 (%)	Precision
All Sites	829	5,591	3,106 ± 385	56 ± 7%	12%
Single Family/HZ1	502	5,943	3,224 ± 458	54 ± 8%	14%
Single Family/HZ2	135	Excluded due to precision greater than 100%			
Single Family/HZ3	0	Excluded due to less than 40 sites (no sites)			
Mnf. Homes/HZ1	160	4,357	3,849 ± 817	88 ± 19%	21%
Mnf. Homes/HZ2	32	Excluded due to less than 40 sites			
Mnf. Homes/HZ3	0	Excluded due to less than 40 sites (no sites)			

Source: Navigant analysis

Through further investigation, the evaluation team found that participants used less heating than expected. The RTF currently provides different UES values for ASHP conversions in homes with "Good," "Fair" and "Poor" insulation. The heating consumption of participants in the sample seemed to better align with the RTF's estimates for homes with "Good" insulation, where the evaluation team together with stakeholders expected participants to have "Fair" insulation. Table E-6 provides additional details.

Table E-6: Participant and RTF's Estimated Energy Use and Savings by Customer Subset

Heating Zone and Home Type	Source	Insulation Type	Heating Load (kWh/year)	Savings (kWh/year)	Percent Savings
All	Billing analysis	All	9,967	3,705	37%
HZ1, SF	RTF estimate	Good	9,383	3,711	40%
		Fair	13,358	6,327	47%
		Poor	19,063	8,943	47%
HZ2, SF	RTF estimate	Good	11,871	3,605	30%
		Fair	16,427	6,098	37%
		Poor	23,739	8,591	36%

Source: Navigant analysis and RTF staff analysis

E.2.2 Performance Duct Sealing

For performance duct, sealing the team found that savings were quite variable and lower than expected. Table E-7 lists the evaluated savings and realization rate for all sites as well as by home type and heat zone. More detailed results are not presented where the sample size was small (i.e., less than 40 sites) or where the precision was unacceptable (i.e., greater than 100%).

Table E-7: Performance Duct Sealing Measures Evaluated Savings by Home Type and Heat Zone

Home Type/Heat Zone	Number of Participants	Savings per Site based on UES Dec. 2016 (kWh)	Evaluated Savings per Site (kWh)	Realization Rate over UES Dec. 2016 (%)	Precision
All Sites	6,187	986	322 ± 148	33 ± 15%	46%
Single Family/HZ1	899	705	551 ± 303	78 ± 43%	55%
Single Family/HZ2	123	Excluded due to precision greater than 100%			
Single Family/HZ3	4	Excluded due to less than 40 sites			
Mnf. Homes/HZ1	3,375	951	275 ± 157	29 ± 16%	57%
Mnf. Homes/HZ2	1,524	Excluded due to precision greater than 100%			
Mnf. Homes/HZ3	262	Excluded due to precision greater than 100%			

Source: Navigant analysis

As listed in Table E-7, duct sealing in single family homes seems to perform closer to expectations than in manufactured homes. No multifamily homes were included in the sample.

E.3 Cost-Effectiveness

Using ProCost and the evaluated savings values, the team estimated cost-effectiveness for ASHP conversions and performance duct sealing. As shown in Table E-8, all measure groups are cost-effective, although the TRC of performance duct sealing is close to 1.0.

Table E-8: Cost-Effectiveness for Residential HVAC UES Measures

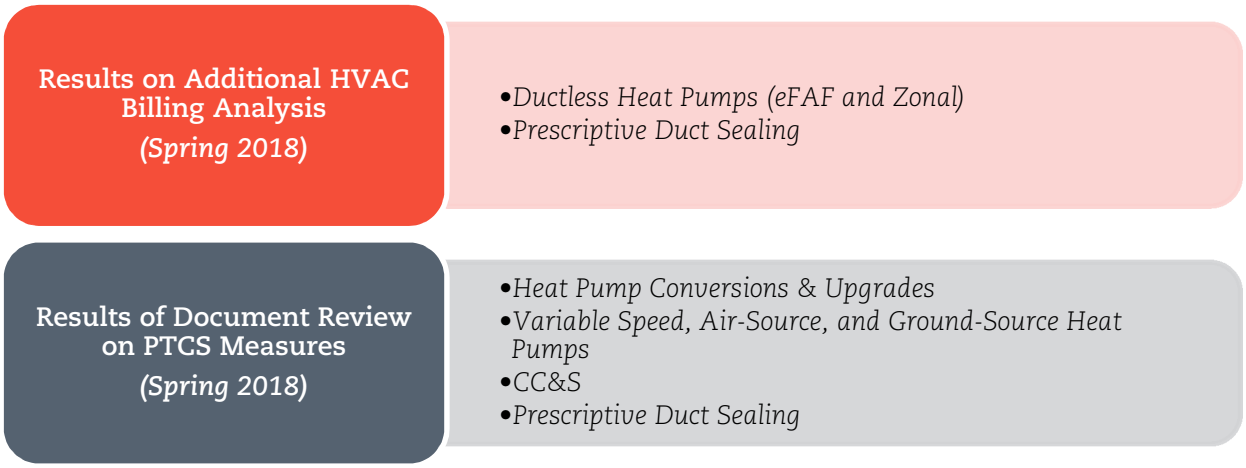
Measure Group	Present Value of Benefits per Unit	Present Value of Costs per Unit	Total Resource Benefit/Cost Ratio
Performance Duct Sealing	$620.28	$509.28	1.22
ASHP Conversions	$7,578.45	$3,898.41	1.94
ASHP Conversions with Duct Sealing	$6,485.11	$4,370.48	1.48

**Non-participant savings are not included in the cost-effectiveness analysis.*
Source: ProCost Analysis using 7th Power plan inputs

E.4 Future Research

The evaluation team recommends letting the results of the on-going research summarized in Figure E-1 inform the need and direction of additional investigation into these measure groups.

Figure E-1: Ongoing Research for Residential HVAC Measures



Appendix F: Billing Analysis Methods

Billing Analysis Methodology

To: BPA

Subject: Methods memo

Date: August 19, 2016

As a part of the 2016 impact evaluation of Bonneville Power Administration's (BPA) Unit Energy Savings (UES) portfolio, Navigant will collect billing data for a representative sample of utilities and use a pooled regression model with a comparison group to estimate measure level savings for select residential HVAC and envelope measures. Navigant's analysis will include robustness checks, data visualization, and statistical tests to validate the results.

This document presents the proposed methods, although minor changes may occur as the project progresses. The following items will be specifically addressed in detail:

- Task 1: Data collection and preparation
- Task 2: Data validation and processing
- Task 3: Primary regression analysis
- Task 4: Secondary analysis
- Task 5: Reporting

F.1 Task 1: Data Collection and Preparation

This section discusses the data collection and preparation methodology for the billing data, weather data, Interim Solution 2.0 data (IS2.0), and the Regional Technical Forum's (RTF's) Unit Energy Savings (UES) data.

F.1.1 Billing Data

To limit the burden on BPA's utility customers, Navigant uses a sampling design that yields a representative sample for the region while minimizing the number of utilities from which Navigant will request billing data. For the sampled utilities, Navigant will request a census of participant billing data (from two years prior and everything following the measure installation) for all evaluation measure groups. The sampled utilities include 22 utilities that are used to represent the region. The evaluation measure groups include four measures:

- Performance and prescriptive duct sealing (performance duct sealing data will come from BPA's already collected data for its evaluation of PTCS measures in 2013)
- Ductless heat pumps (DHP) replacing electric forced air furnaces (eFAF)
- Windows
- Insulation

Details of the sample designs for the residential envelope and HVAC domains are provided in Table F-1 and **Error! Reference source not found.**, and additional information can be found in Sections 5 and 6 of the Bonneville Power Administration UES Portfolio Evaluation Plan CY2016 Activities document.[12]

Table F-1: Draft 2016 Sample Size for the Residential Envelope Domain

Measure Group	Strata	Assumed CV	Number of Utilities	Target Number of Projects*	Confidence and Precision Targets
Insulation	Large Contributors	0.8	4	Census of billing data for each utility**	90/15
	Medium Contributors	0.8	4		
	Small Contributors	0.8	2		
	Subtotal		**10**	**~1,700**	
Windows	Large Contributors	0.8	2	Census of billing data for each utility**	90/15
	Medium Contributors	0.8	10		
	Small Contributors	0.8	3		
	Subtotal		**15**	**~2,500**	

* This value represents the target number of projects for which the evaluation team requires usable data. In order to reach this number, the team will need to request billing data for roughly twice as many projects.

** Evaluation will target a census of energy consumption data for sampled utilities; a sample may be drawn where this is infeasible.

Source: Navigant Analysis

[12] https://www.bpa.gov/EE/Utility/research-archive/Documents/Evaluation/BPA_UES_Evaluation_Plan_FINAL_04012016_V3.pdf

Table F-2: Draft 2016 Sample Size for the Residential HVAC Domain

Measure Group	Strata	Assumed CV	Number of Utilities	Target Number of Projects*	Confidence and Precision Targets
Prescriptive Duct Sealing	Large Contributors	0.8	5	Census of participant billing data**	90/15
	Medium Contributors	0.8	3	Census of participant billing data**	
	Small Contributors	0.8	2	Census of participant billing data**	
	Subtotal		**10**	**~500**	
Ductless Heat Pumps replacing Forced Air Furnaces	Large Contributors	0.8	6	Census of participant billing data**	90/15
	Medium Contributors	0.8	6	Census of participant billing data**	
	Medium and Small Contributors	0.8	2	Census of participant billing data**	
	Subtotal		**14**	**~800**	

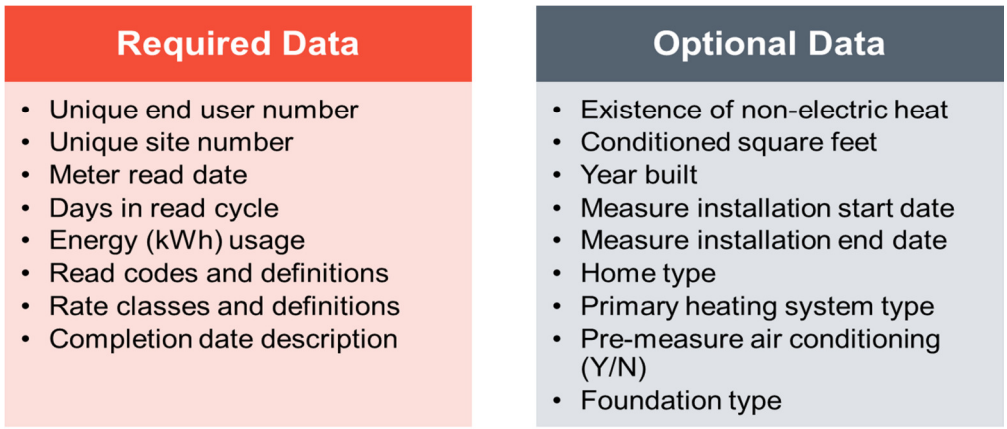
* This value represents the target number of projects for which the evaluation team requires usable data. In order to reach this number, the team will need to request billing data for roughly twice as many projects.

**Navigant does not estimate sampling error or a sample size for these measure groups, because the evaluation team plans to collect data on the census of these projects within the sampled utilities.

Source: Navigant Analysis

The data Navigant will request includes the required and optional data listed in Figure F-1. Navigant will request this data from each of the 22 utilities using a standardized data template in Excel.

Figure F-1: Required and Optional Data Requested from Utilities



Source: Navigant

F.1.2 Weather Data

Navigant will gather two sets of weather data (actual and typical year) that will be used to control for varying weather conditions in the analysis and to calculate weather-normalized savings (i.e., savings during a typical meteorological year (TMY)). Navigant will use the study participants' zip codes to match their locations to weather stations in the National Oceanic and Atmospheric Administration's National Climatic Data Center[13] database of historical weather data. Navigant will ensure that each zip code provided by the utilities is reasonable given each utility's service territory. Navigant will also ensure that each matched weather station has adequate weather data and provides TMY data from the National Renewable Energy Lab's National Solar Radiation Database[14]. The analysis requires historical weather data during the study period and in the study location to properly correlate participants' historical energy consumption with weather conditions. TMY data is then used to adjust savings to show expected savings for typical years.

F.1.3 IS2.0 Data

BPA's Interim Solution 2.0 database (IS2.0) includes all measures incentivized by and reported to BPA. The IS2.0 database will be queried to identify all efficiency measures installed at study participant sites during the evaluation period and when those measures were installed. Navigant will use this data to ensure their billing analysis does not double count savings and to ensure that their analysis accounts for real world measure interaction. Double counting in this context refers to attributing all energy savings at a site, which often come from multiple measures, to a single measure. Knowing all the measures installed at each study participant site enables Navigant to attribute savings to single measures more accurately. Navigant uses the term "measure interaction" to refer to the diminishing savings from installing multiple measures at a

[13] http://www.ncdc.noaa.gov/cdo-web/
[14] http://rredc.nrel.gov/solar/old_data/nsrdb/1991-2005/tmy3/

single site. For example, converting a home from an electric forced air furnace to a heat pump increases the heating efficiency of that home. Installing attic insulation at the same time now yields lower savings attributable to insulation than the insulation would have achieved if the site hadn't converted to a heat pump. The overall site-level savings is higher, but the savings attributable to insulation is likely lower when installed simultaneously with a HP conversion.

F.1.4 Regional Technical Forum's Unit Energy Savings Measures Data

Navigant will use the Regional Technical Forum's (RTF's) Unit Energy Savings (UES) for the following points of comparison against the regression estimates of savings from this billing analysis:

- **Claimed savings –** the UES at the time of installation and associated with the study participants will serve to answer the question, *did these efficiency measures save as much energy as originally anticipated on average?*
- **Most recent savings –** the most recent UES associated with the relevant measures will serve to answer the question, *do the savings from previous participants corroborate the most recent UES proposed by the RTF to track savings for future participants for these measures on average?*

F.2 Task 2: Data Validation and Processing

This section discusses the methods Navigant will use to process and validate the requested billing data. Navigant will track the learnings from these analyses with additional variables in the dataset rather than removing the data outright. This approach will enable future investigations of the effect on savings from different data filtering strategies. Navigant describes the proposed approach in the following subsections:

- Billing data clarification and processing
- Utility, BPA, and weather data – validation and summary
- Initial outlier analysis to further understand dataset
- Data filtering

F.2.1 Task 2a: Billing Data Clarification and Processing

After receiving the requested billing data, Navigant will conduct standard data quality checks and follow up with the sampled utilities with any questions. These follow up questions will ensure that Navigant interprets the data accurately and understands the program approach, which may influence some analyses. Navigant will check for duplicate meter reads, negative meter reads, short or long billing periods, unexpected read codes, low and high energy consumption, overlapping billing periods, and other standard data quality investigations. Navigant outlines each check in greater detail in

F.6.2, along with examples of the outputs used to identify these issues. When data anomalies are identified, the utility will be contacted so the evaluation team can determine the cause of the anomaly (e.g., if it is due to erroneous data, misinterpretation of the data, etc.) and to ensure Navigant uses the data accurately.

Finally, Navigant will process the data (including any adjustments based on the responses to the follow up questions) so it is ready to be used in the billing analysis.

F.2.2 Task 2b: Utility, BPA, and Weather Data – Validation and Summary

After Navigant conducts standard data quality checks on the received billing data, including any follow up with the sampled utilities, Navigant will compile the billing data, weather data, and IS2.0 data. Navigant will validate this data at a high-level with the following summary checks:

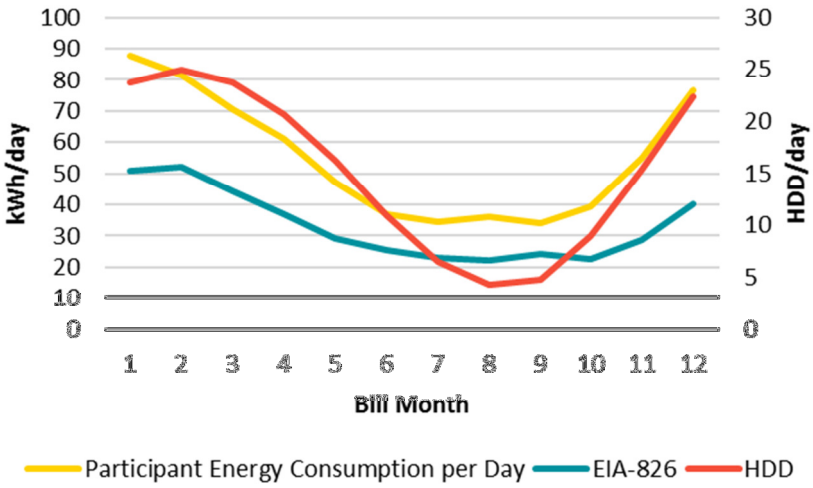
1. Summarize sample size by home type and measure.
2. Summarize energy consumption, heating degree days (HDD), and cooling degree days (CDD) by month and year for pre-and post-installation data.
3. Compare participants' energy data by bill month to weather data and EIA-826[15] data (see Figure F-2 for an example) to verify the received data is the expected order of magnitude.

An example of the graph Navigant produced during a preliminary analysis for the third summary step is provided in Figure F-2. In this graph, participant energy consumption is based on billing data and is shown in red as average daily energy consumption (kWh/day) for each calendar month. The grey line corresponds to the average daily energy consumption per household for each calendar month based on the EIA-826 data (kWh/day). The weather data is represented by the yellow line and shows average HDD per day for this anonymous utility territory.

This graph allowed Navigant to compare data from different sources, identify expected trends and unexpected discrepancies, and ultimately to validate that no systematic errors occurred in collecting this data. The difference in participant and EIA-826 energy consumption is expected, because the participant group primarily consists of electric heat customers while the EIA-826 data has a much higher penetration of non-electrically heated customers.

[15] The EIA-826 data comes from a database provided by the Energy Information Administration (EIA) that includes publicly available residential retail electricity sales by utility and month, and the number of customers in each utility

Figure F-2: Validation of Energy and Weather Data



Source: Navigant Analysis

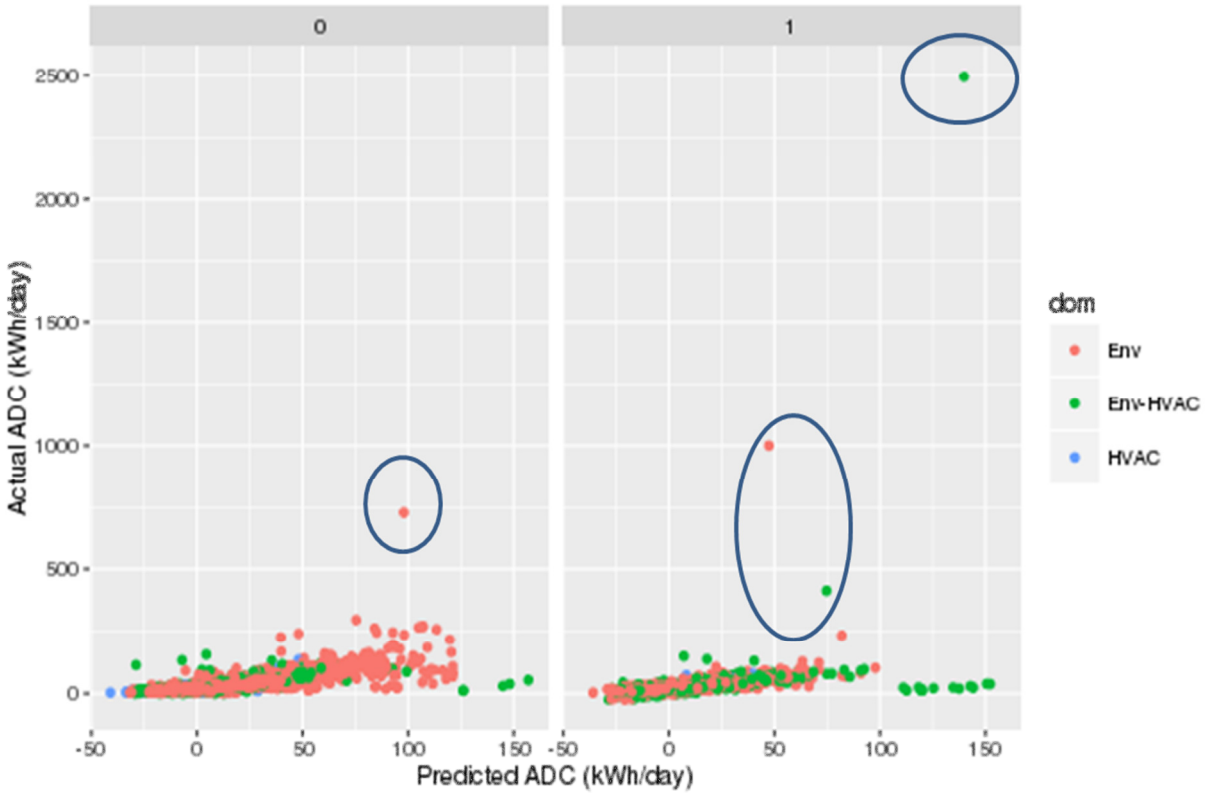
F.2.3 Task 2c: Initial Outlier Analysis to Further Understand Dataset

Navigant will conduct an initial outlier analysis to further refine the dataset by identifying any sites or bills that require additional investigation. Navigant will identify outliers by graphing participants' actual average energy consumption against their modeled energy consumption using a simplified econometric model based on site specific energy consumption and weather conditions. Navigant will identify outliers with a visual approach rather than a quantitative method for identifying outliers. When outliers are identified, Navigant will review the available data at these sites and bills, and propose methods to resolve any discovered issues. Possible issues to resolve are likely to include:

- Poorly estimated bills with substantial corrections
- Unexpected read codes
- Short bill periods (e.g., 1 day)
- High prevalence of bills with 0 kWh usage

Figure F-3 shows an example of the outlier analysis conducted during a preliminary analysis for this project for an anonymous utility. Additional investigation of the five bills circled in blue enabled Navigant to justify removing these five bills from the analysis.

Figure F-3: Average Daily Consumption Outliers



Source: Navigant Analysis

F.2.4 Task 2d: Data Filtering

After identifying all discovered anomalies, Navigant will track potential causes for data removal using additional variables in the dataset rather than removing the data outright. This approach will enable future investigations of data filtering's effect on savings estimates. Example data filters could include the following and please see the table in F.6.6 for proposed data filters for the primary regression analysis.

- Minimum number of bills before and after measure installation
- Verifiable heat signature (before and after measure installation) using a site-specific R-squared threshold
- Verifiable heat signature (before and after measure installation) using a site-specific root mean squared error (RMSE)
- Quality comparison group match (when using Euclidean matching) as measured by RMSE
- Bills deemed erroneous due to read code or magnitude of consumption (e.g., 10 times the median)
- Bills indicating vacancy

- Homes with an indication of net metering (e.g., through rate class)
- No longer relevant applications of a measure
- Homes experiencing very large increases or decreases in consumption (e.g., 50% or more)

F.3 Task 3: Primary Regression Analysis

Navigant will use a pooled regression model with a comparison group to estimate measure level savings for the relevant HVAC and envelope measures.

Navigant will compare estimated savings from the billing analysis to both claimed UES values, and to the most recent UES values available in the RTF's interim workbooks. The evaluation team will present these results as absolute savings and as realization rates (RR), where:

$$RR = \frac{Regression\ estimate\ of\ savings\ \left(\frac{kWh}{yr}\right)}{Claimed\ or\ most\ recent\ UES\ estimate\ of\ savings\ \left(\frac{kWh}{yr}\right)}$$

Realization rates are a useful comparison because they help account for the diversity among participants and measures. In addition, comparing the savings estimates with existing deemed values will help identify measures that may need updates or further research.

The evaluation team will also use these savings results to estimate the cost-effectiveness of these measures. In the subsections below, Navigant discusses the following in further detail.

- Robustness checks
- Comparison groups
- Regression models
- Weather normalization

F.3.1 Task 3a: Robustness Checks

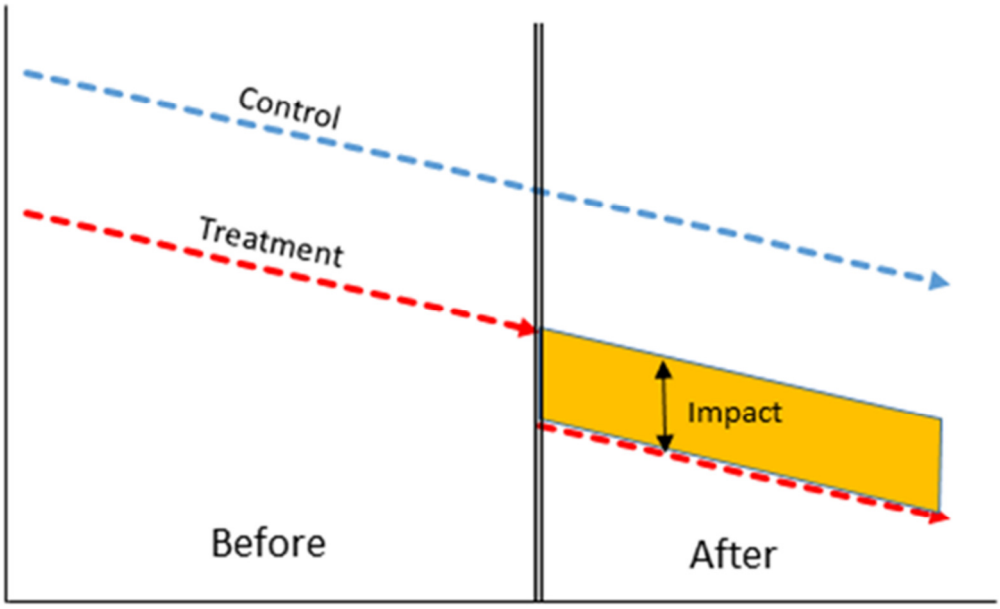
Due to unverifiable assumptions inherent in any savings approach, Navigant proposes to conduct a recommended approach to estimate savings as well as multiple alternate approaches to serve as robustness checks. Navigant uses the term "robustness check" to refer to savings estimated through alternate methods that verify the robustness of the results to the relevant assumptions. Each method has a different set of assumptions. If the savings are similar across these checks, it is likely that any assumptions are held true and the results can be considered accurate. If the results differ, some assumptions may not hold true. In advance of finding inconsistent results, Navigant will propose their preferred approach in this document to avoid any

perceived "cherry-picking" of results.[16] Only the final and recommended results are shared in the final report to avoid confusion, but BPA can provide additional details upon request.

F.3.2 Task 3b: Comparison Group

Navigant will use a comparison group (in this case pre-measure-installation electric usage from future participants) to isolate the effects of the program on energy consumption and to exclude any non-program related changes in energy consumption, such as changes in energy consumption due to economic conditions. Figure F-4 provides a conceptual diagram to demonstrate the value of a comparison group. Furthermore, in preliminary analyses for BPA, Navigant found that accounting for non-program related changes in energy consumption had a statistically significant effect on the estimate of savings (Figure F-6).

Figure F-4: Conceptual Diagram of the Value from a Comparison (or Control) Group



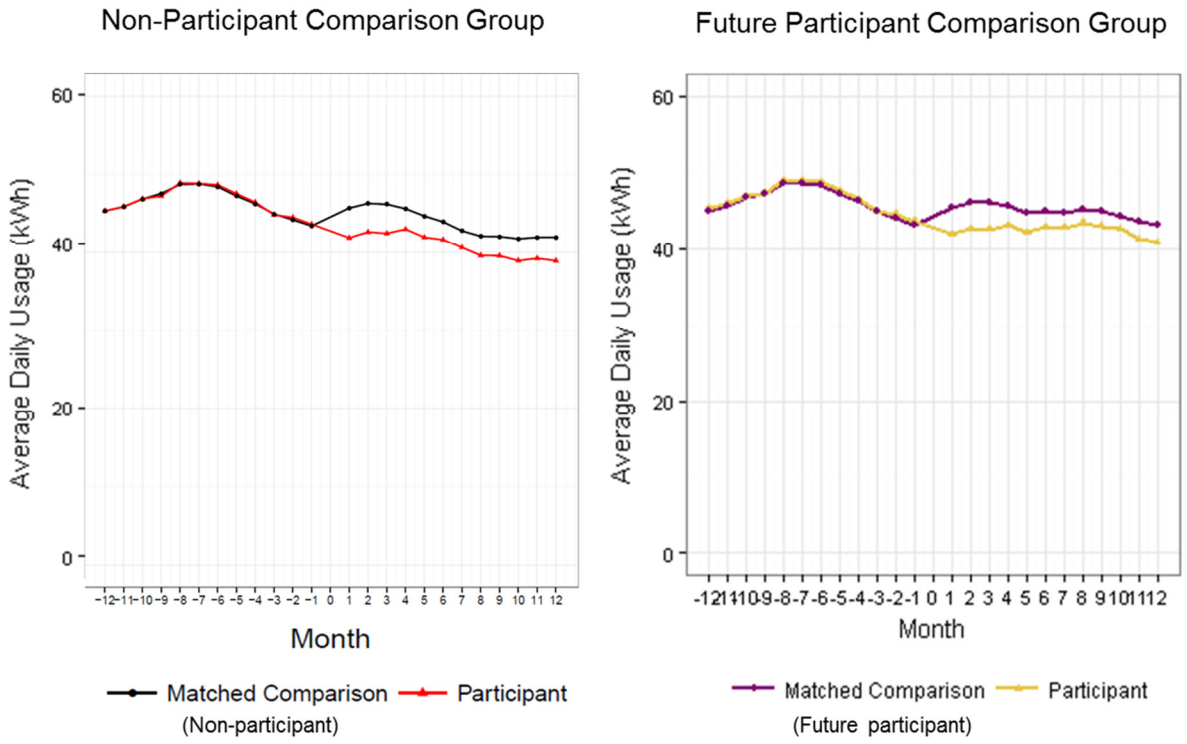
Source: Navigant

While randomized experimental designs are the ideal approach to account for non-program related changes in energy consumption, that type of study design is not feasible at this time (see F.6.5 for more information). Instead of conducting a randomized experiment, Navigant develops a comparison group by selecting sites from future participants' pre-measure-installation energy usage (also referred to as "pre-usage"), such that the comparison group and evaluated participants share similar pre-usage. Navigant supports this approach as an alternate to an experimental design for the following reasons:

[16] It is important to note that all of the proposed methods are subject to self-selection bias, which can only be addressed using an experimental design. Please refer to F.6.5 for more information on this topic.

- To reduce the burden on utilities – implementing an experimental design or requesting non-participant energy consumption with 20+ utilities would put substantial burden on the utilities running these programs
- Recommended in Uniform Methods Project (UMP)[17] – using future participant pre-usage is a recommended approach in the UMP
- Reasonable results in preliminary analysis – in preliminary analyses, non-participants and future participants performed similarly when used as comparison group sources or reservoirs (Figure F-5 and Figure F-6). Figure F-6 shows that savings were not statistically different when non-participants and future participants (orange and green columns) served as the comparison group reservoirs, but savings were statistically different when no comparison group was used (red column). It was only possible to collect non-participant data for study 1 and, as such, there is no green column.

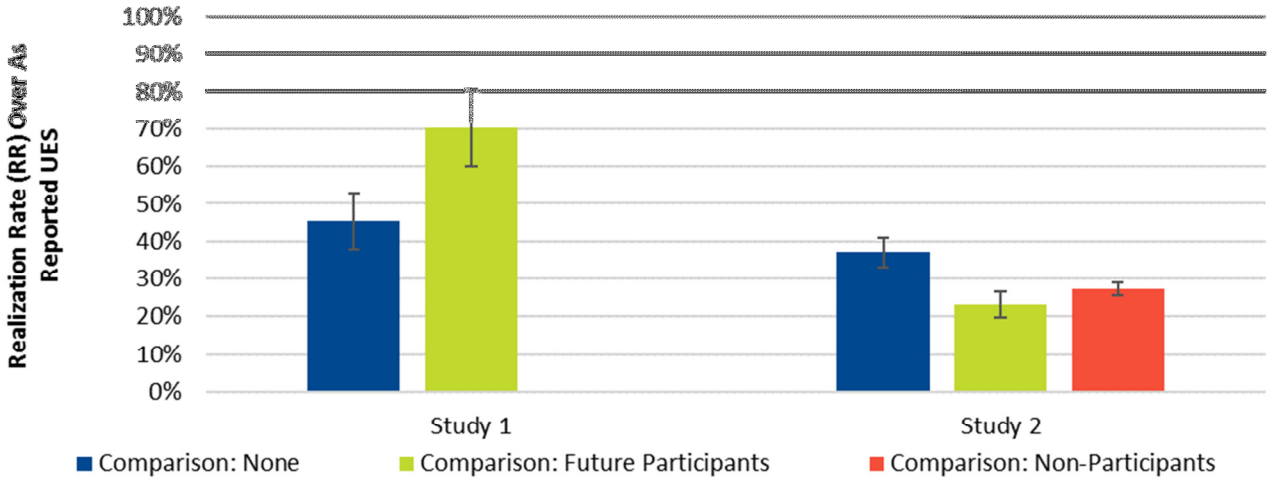
Figure F-5: Pre and Post Energy Usage for Participants and Various Comparison Groups



Source: Navigant Analysis

[17] http://energy.gov/sites/prod/files/2013/11/f5/53827-8.pdf

Figure F-6: Average Realization Rates Across Measures



Source: Navigant Analysis

For this study, Navigant will validate the comparison group by using graphing techniques (as demonstrated in Figure F-5 and Figure F-6) and statistical tests for all proposed methods.

F.3.3 Task 3c: Estimating Savings

This section will describe the following methods to estimate savings in more detail:[18]

- Post-only model with Euclidean matching (recommended approach)
- Variation-in-adoption model (robustness check)
- Difference-in-difference model (robustness check)
- Variable based degree days (robustness check)

Navigant chooses the post only model as the recommended approach because of its demonstrated success in randomized control trial (RCT) evaluations (e.g., Opower evaluations) and its ability to control for unobservable, non-program related changes in energy consumption, such as economic changes.

F.3.3.1 Post-Only Model with Euclidean Matching (Recommended Approach)

This model uses pre-usage as an independent variable to predict post-installation energy usage (also referred to as "post-usage"). This model gets its name because pre-usage shows up on the right-hand side of the regression equation (as an independent variable) rather than on the left-hand side. Savings then corresponds, in a simplified

[18] Navigant will also estimate savings using a site fixed effects model without a comparison group as an additional point of reference with the intention of understanding the impact of a comparison group on the savings estimate.

interpretation, to the difference in energy consumption between the participants' actual and predicted post-usage. Controlling for monthly fixed effects and having a balanced comparison group and participant group ensures that the model's estimated savings refers only to the change in energy consumption due to the program and not from other effects, such as economic conditions.[19] Navigant discusses this method in greater detail in F.6.4, including a basic model specification.

Navigant ensures that the groups are balanced by using exact matching based on pre-usage, heat zone, and home type. Each matched comparison group site receives a pseudo measure installation date from its matched participant site. In summary, exact matching selects sites from the comparison group reservoir that best represents participants. More details on this comparison group approach are provided in F.6.3.

In summary, the model:

- Estimates average post-usage of customers as a function of pre-usage, monthly fixed effects and other factors.
- Estimates savings as the impact on post-usage of being a participant (as opposed to being one of the matched comparison sites) or, in a sense, the difference between participants' predicted and actual post-usage.

F.3.3.2 Variation-in-Adoption Model (Robustness Check)

The VIA method relies only on program participants to develop the counterfactual[20] (i.e., the energy participants would have consumed had they not participated in the program). Customers who participate in the program at a later date serve as the comparison group for the customers who participate in the program early on. The main assumption of the VIA model is that after controlling for both customer and monthly fixed effects neither average energy use in a month nor energy savings each month after participation is correlated with the timing of program entry, or in other words program participant savings and consumption are similar regardless of when customers participate. For example, weather effects on energy consumption would violate this assumption (consumption would be higher after participating in October as heating systems turn on) without controlling for weather using monthly fixed effects. The model tests this assumption with specific independent variables that should ideally have not statistically significant coefficients. A basic model specification is provided in F.6.4.

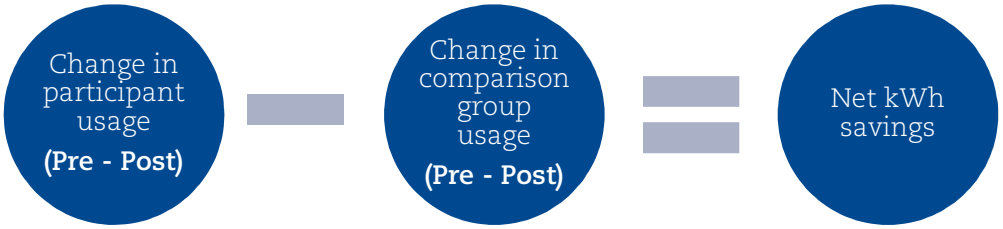
[19] Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth Stuart. 2007. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis* 15(3): 199-236.

[20] Description of VIA starts on page 16 of Harding, Matthew and Hsiaw, Alice, "Goal Setting and Energy Conservation" (2014). Economics Department Working Papers. Paper 166. http://crossworks.holycross.edu/econ_working_papers/166

F.3.3.3 *Difference-in-Difference Model (Robustness Check)*

The DID model estimates savings based on the change in energy consumption at the time of participation for the participants *minus* the change in consumption at the time participants participated for the comparison group (Figure F-7).

Figure F-7: Illustration of DID Model



Source: Navigant

For this study, Navigant will use FY2014 participants as the "participants" and FY2015 participants' pre-usage as the comparison group. Constructing comparison groups based on future participants is a recommended approach in the UMP.[21] A basic model specification and a more detailed discussion of DID is provided in F.6.4. In summary, the DID model:

- Calculates the average difference in consumption between the pre-and post-periods for the participants.
- Calculates the average difference in consumption between the pre-and post-periods (as defined by the participants' measure installation dates) for the comparison group.
- Subtracts these two differences to estimate the net energy savings (Figure F-7 illustrates the net savings for a two-period model).

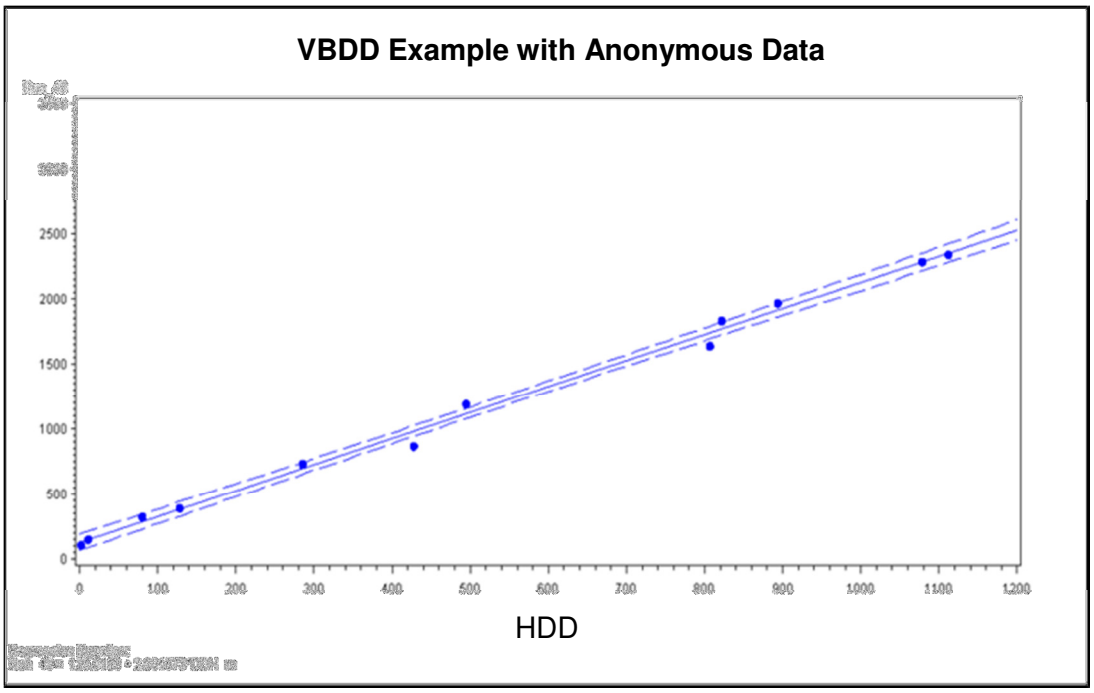
F.3.3.4 *Variable Base Degree Days (Robustness Check)*

The Variable Base Degree Day (VBDD) method fits a model that reflects the specific energy consumption dynamics for each site's pre-and post-installation energy consumption one at a time. This approach optimizes the balance temperature (for HDD and CDD) for each model.[22] Figure F-8 provides an example of a VBDD model fit to just HDD with anonymous data.

The evaluation team will calculate savings with a VBDD approach to serve as an additional robustness check. Navigant will compare its VBDD results to the VBDD results from an open-source package developed by Ecotope and RTF staff to ensure there are no substantial deviations in methodology.

[21] http://energy.gov/sites/prod/files/2013/11/f5/53827-8.pdf

Figure F-8: Example of VBDD



Source: Navigant

F.3.4 Task 3d: Weather Normalizing Savings

Navigant will estimate the weather-normalized average measure-level energy savings (kWh) by interacting HDD and CDD with the savings term. In the dataset, HDD and CDD will reflect actual weather conditions. Their corresponding coefficients will then be applied to TMY HDD and CDD to reflect typical weather year savings.

The model specification below (Equation F-1) provides a conceptual example of a post-only model that can yield weather normal savings.

Equation F-1: Post-Only Regression Model with Weather Adjustment

$$ADC_{kt} = \sum_{J} \beta_{1j} Month_{jt} + \sum_{J} \beta_{2j} Month_{jt} \cdot ADClag_{kt} +$$
$$\beta_3 Participant_k \cdot HDD_t + \beta_4 Participant_k \cdot CDD_k + \varepsilon_{kt}$$

Where

ADC_{kt} = The average daily usage in kWh for customer *k* during billing cycle *t*. This is the dependent variable in the model

$Month_{jt}$ = A binary variable taking a value of 1 when j=t and 0 otherwise[23]

$ADClag_{kt}$ = Customer k's energy use in the same calendar month of the pre-program year as the calendar month of t

$Participant_k$ = A binary variable indicating whether customer k is in the participant group (taking a value of 1) or in the control group (taking a value of 0)

HDD_t = Heating degree days in a given month t

CDD_t = Cooling degree days in a given month t

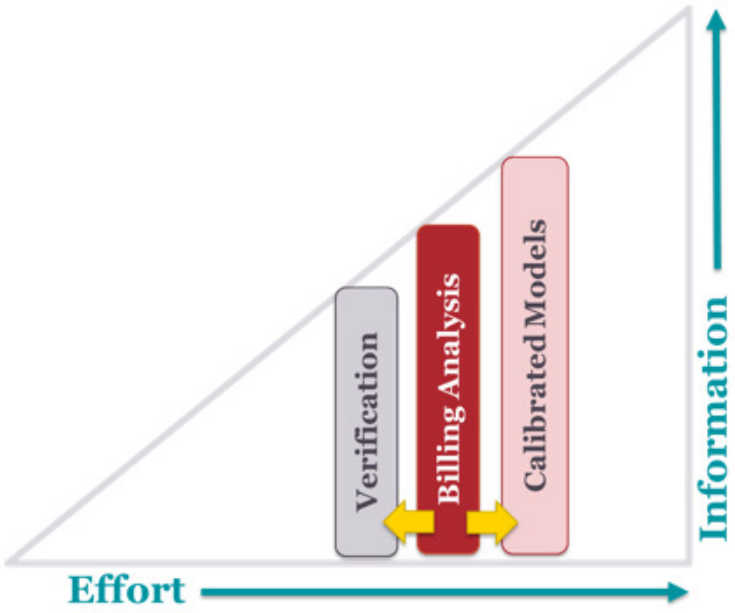
ε_{kt} = The cluster-robust error term for customer k during billing cycle t. Cluster-robust errors account for heteroscedasticity and autocorrelation at the customer level

F.4 Task 4: Secondary Analysis

In this step of the project the evaluation team will attempt to provide more information on the savings to (1) help explain why the savings are at their current level and (2) further understand the impact of analysis decisions on results. As shown in Figure F-9 increasing the level of effort and investigation for this project can increase the information provided to BPA and its relevant stakeholders, allowing BPA and its stakeholders to reach more informative conclusions. Delivery verification can reveal discrepancies between the number of measures tracked as being installed and the actual number installed, a billing analysis can estimate the real-world savings from the installed measures, and additional investigation through phone surveys, calibrating models, etc. can help explain why the savings are at their current level.

[23] If there are T post-program months, there are T monthly dummy variables in the model, with the dummy variable $Month_{tt}$ the only one to take a value of 1 at time t. These are, in other words, monthly fixed effects.

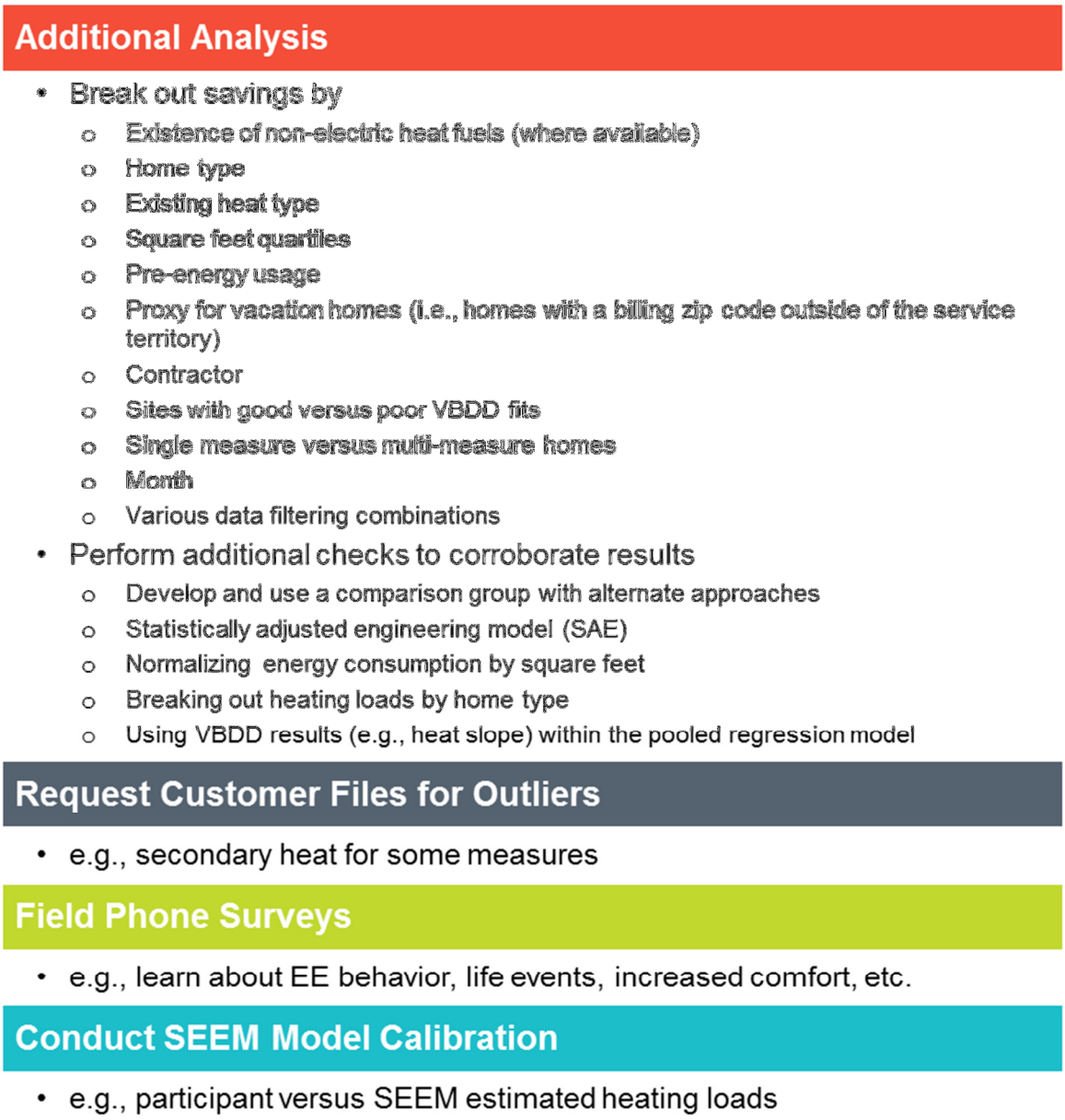
Figure F-9: Increased Effort Can Lead to More Informative Conclusions



Source: Navigant

This analysis may include some of the components outlined in Figure F-10 and the evaluation team will work with BPA and its stakeholders to determine both the initiation, type of analysis and timing of the second stage approach. The analysis will at least run 10 additional models with the already collected data, where BPA and relevant stakeholders will provide guidance for the models to run, and compare total consumption between SEEM outputs and the collected billing data. If the billing analysis results are significantly different than the RTF best-available savings estimates, the evaluation team may develop an analysis plan to further explain the savings results. Staging this second piece of analysis will allow BPA and its stakeholders to minimize burden on sampled customer utilities, while still providing the opportunity to gain insight into why estimated savings may vary from claimed savings.

Figure F-10: Components of Navigant's Optional Second Stage Outlier Analysis



Source: Navigant

F.5 Task 5: Reporting

This section discusses Navigant's final deliverables and the schedule to complete the billing analysis.

F.5.1 Deliverables

Navigant's work will result in five final deliverables:

1. Report, which will include a discussion of data filtering and the final model specifications used in the analysis
2. Final savings results and detailed modeling results (by coefficient) in Excel spreadsheet format
3. One-hour webinar with slide deck
4. Highlights document (i.e., one-pager) with high-level results
5. Analysis dataset (including some flexibility for alternate data filtering options)

F.5.2 Schedule

Table F-3 shows the task and deliverables schedule to complete Navigant's billing analysis of residential HVAC and envelope measures.

Table F-3: Project Task and Deliverable Schedule

Task	Approximate Completion Date
Task 1: Data collection and preparation	August 15, 2016
Task 2: Data validation and processing	October 3, 2016
Task 3: Primary regression analysis	January 15,2017
Task 4: Secondary analysis (w/o additional data collection)	February 1,2017
Task 5: Final report, presentation, results, data set and highlights document	March 1, 2017
Task 4: Secondary analysis (w/ additional data collection)	Ongoing

Source: Navigant

F.6 Billing Analysis Methods Appendices

F.6.1 Data Collection Processes

Navigant will support the data collection by using an efficient and organized system, and will provide a staff member who is friendly and helpful to be the primary point of contact for questions on the evaluation and data collection process. The evaluation team will also maintain a log of communication with utility staff and maintain a record of the status of the data collection process with each utility. Navigant will offer multiple methods for data collection and transfer, including a secure file transfer protocol, mailing flash drives or paper documents, and going on-site to support data collection. Navigant may be requested by individual utilities to negotiate a non-disclosure agreement for data privacy.

Navigant realizes that some utilities may have difficulty pulling the requested number of billing records, i.e. for small utilities with large participation. In these cases, the evaluation team will work closely with BPA staff to either assist with billing data extraction, set a lower target than the census, or select a replacement utility as necessary. For large contributors, the team needs to be especially diligent in getting as many participants as possible. Additionally, BPA will request volunteers for utilities outside of the sample to increase sample sizes.

Navigant may also request a small sample of utility customer files to support the evaluation approach of the residential envelope and HVAC domains. Although it would be ideal to receive customer files for all projects in the sample, the evaluation team realizes that this could represent a significant burden to utilities. Instead, at the initial sample stage, Navigant will accept customer project files from utilities who voluntarily agree to provide this data (e.g., those that have an easy system for pulling this information). However, Navigant may request a sub-sample of customer files for select outlier projects after completing the second stage outlier analysis. The evaluation team will request project files at that time. Additionally, the evaluation team will work individually with utilities to support the easy provision of customer files.

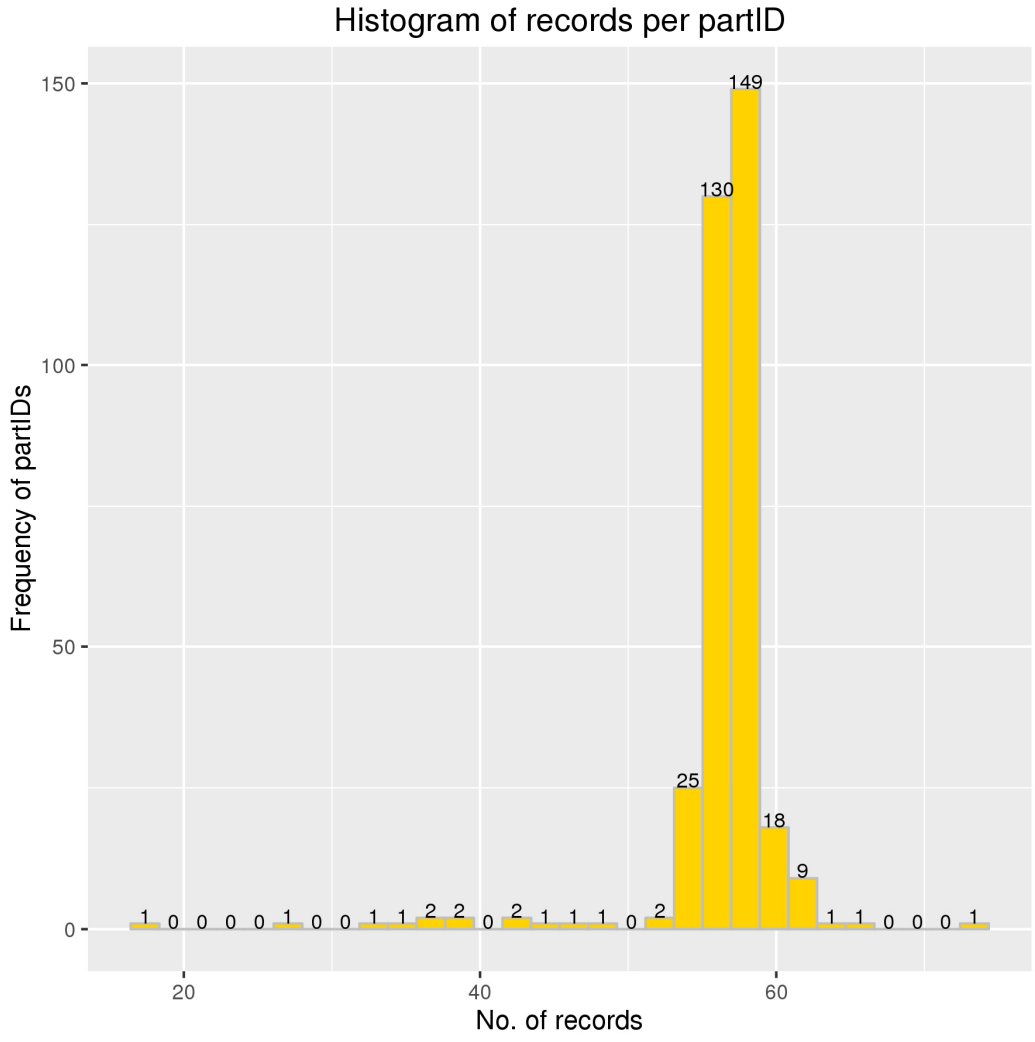
F.6.2 Initial Data Quality Control

Navigant will initially explore the data received from each utility to ensure the quality of the data is adequate to complete the remainder of the analysis. Navigant provides examples of our quality control checks below, by topic.

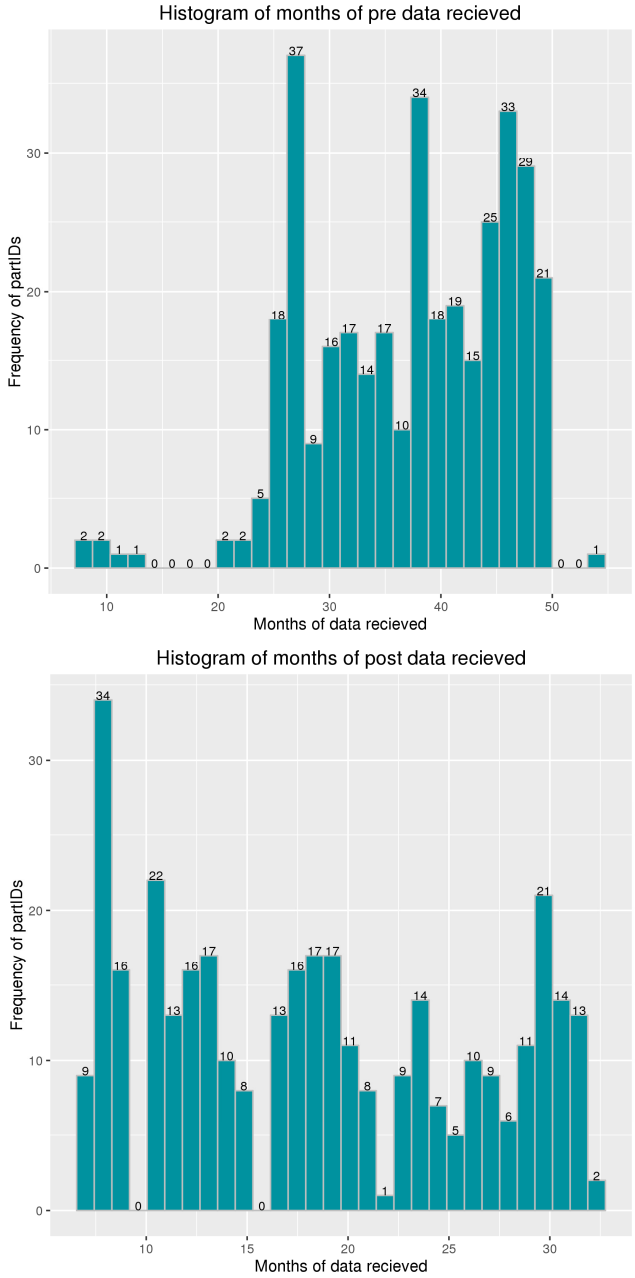
1. To verify that Navigant received the requested data:

Number of Sites Requested	Number of Sites Received
350	349

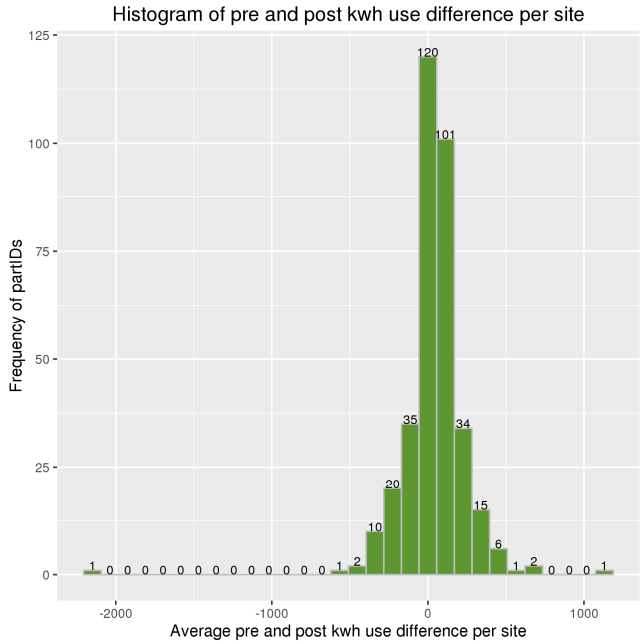
2. To review the number of observations per site:



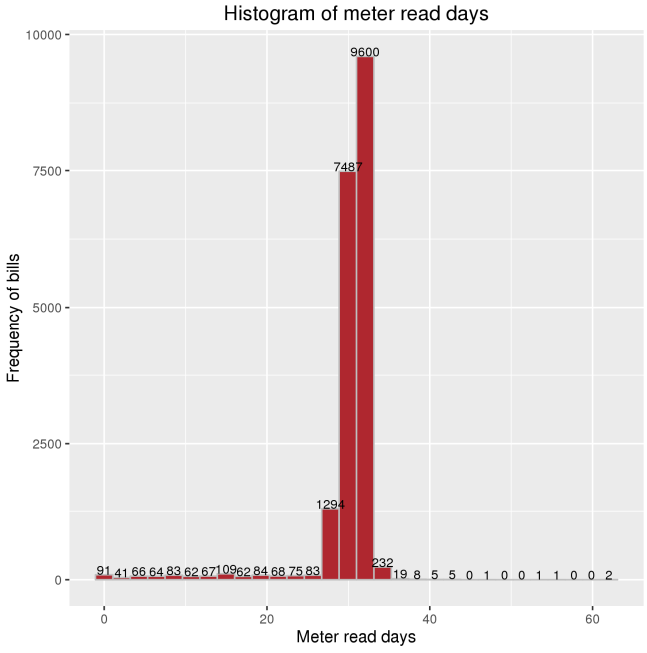
3. To determine the number of sites for which there are enough bills for billing analysis:



4. To determine the range of the change in energy consumption before and after the measure installation:



5. To determine the range of read days included in each meter read:



F.6.3 Detailed Discussion of Comparison Group Approach

A perfect comparison group reflects what participants would have looked like without the program (i.e., the counterfactual). In this case, a perfect comparison group would reflect participants' energy use had they not participated in the program. The effect of the program is then the difference between the average difference in consumption between pre-and post-periods for the treatment group and the average difference in consumption between the pre-and post-periods for the comparison group:

$$\text{Impact} = (Treat_{post} - Treat_{pre}) - (Comparison_{post} - Comparison_{pre})$$

In practice, this is very difficult to accomplish without the experimental design. Thus, Navigant will employ the quasi-experimental design approaches listed below to construct a comparison group using a sample of "future-participants'" pre-usage and validate whether the comparison group is appropriate for the regression analysis to estimate savings from the program. Any non-experimental design is imperfect, and Navigant discusses this topic in greater detail in F.6.5.

- Exact or Euclidean distance matching
- VIA
- Future participants without matching, refinement, or sub-setting

Exact or Euclidean Distance Matching (Matching on Customer Pre-Usage)

This matching method attempts to simulate random sampling of treatment and comparison groups by matching each treatment customer with a comparison group "best match" based on the customer pre-usage. This matching method is widely accepted as a reasonable alternative method to construct a comparison group when an experimental design is not an option.[24]

Matching methods rely on constructing a set of matched comparison households to estimate program savings.[25] The basis of the comparison is the difference in monthly energy use between a participant and a potential match, D_{PM} (**D**ifference between **P**articipant and potential **M**atch). The quality of a match is denoted by the Euclidean distance to the participant over the values of monthly D_{PM} used for matching. In this case the D_{PM} refers to monthly pre-usage and Euclidean distance is measured as the root mean squared error between a given participant's and potential match's pre-usage over the same time period. The non-participant customer with the shortest Euclidean distance to a participant is chosen as the matched comparison for the participant. Navigant may use certain filters for this study, including heat zone and home type, where single family homes can only match to other single-family homes and homes in

[24] State and Local Energy Efficiency Action Network. 2012. Evaluation, Measurement and Verification (EM&V) of Residential Behavior-Based Energy Efficiency Programs: Issues and Recommendations. Prepared by A. Todd, E. Stuart, S. Schiller, and C. Goldman, Lawrence Berkeley National Laboratory.
[25] Alcott, H, and T. Rodgers. "The Short-run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." American Economic Review, 104(10): 3003-37. 2014.

heat zone 1 can only match to other homes in heat zone 1, etc. Matching for this study will be done with replacement, and the standard error will account for this by using a robust standard error that clusters the error around the individual at every instance of each individual.

Navigant will use "future participants" (i.e., future participants' pre-usage) as opposed to non-participants to construct the comparison group. More specifically, the basic logic of matching is to balance the participant and future-participant samples by matching on the exogenous covariates known to have a high correlation with the outcome variable, which is monthly post-program period energy usage. Doing so increases the efficiency of the estimate and reduces the potential for model specification bias. The fundamental assumption on matching is that the distribution of the observed covariates ("pre-usage") is the same for treatment and control observations, which reduces potential model specification bias while gaining statistical power to estimate savings. In practice Navigant finds that pre-usage is very highly correlated with post-usage.

Variation-in-Adoption Model

Controlling for customer and time fixed-effects, the VIA model relies on the assumption that neither energy use in month *t*, nor energy savings *s* months into the program, is correlated with the timing of program entry. In a sense, this model takes advantage of other participants' consumption in a program with rolling participation. Within the model it is common to use independent variables that check this assumption. Navigant will use the VIA analysis as a robustness check for the matching effort.

Future Participants without Matching, Refinement, or Sub-Setting

When non-participant billing data is not available to construct a comparison group, future participants (without matching) may serve as a representative comparison group and may provide an adequate estimate of the counterfactual for the evaluated participants within a DID model framework. This comparison group will serve as a robustness check for the matching effort, but is considered less preferable as there is no effort to ensure the comparison group is representative of the evaluated participants.

F.6.4 Basic Regression Model Specifications

In this section Navigant provides basic model specifications and detailed descriptions for the post-only, VIA, and DID models.

Post-Only Model

The post-only model assures that the distributions of the explanatory variables (e.g., average daily energy consumption) for participants are the same as those for the comparison group. In this approach the development of a matched comparison group is viewed as a useful pre-processing step in a regression analysis to assure that the distributions of the covariates (i.e., the explanatory variables on which the output variable depends) for the treatment group are the same as those for the comparison group. This minimizes the possibility of model specification bias.

The post-only model combines both cross-sectional and time series data in a panel dataset. This model uses the post-program data for the dependent variable (average daily energy consumption) and uses lagged energy use for the same calendar month of the pre-program period to control for any small systematic differences between the participant and control customers. Energy use in calendar month *t* of the post-program period is a function of both the participant variable and energy use in the same calendar month of the pre-program period. The underlying logic is that systematic differences between participants and controls will be reflected in differences in their past energy use, which are highly correlated with their current energy use. Navigant includes an interaction term of pre-program energy use and monthly fixed effects to allow pre-program usage to have a different effect on post-program usage in each calendar month. A basic, conceptual model specification is provided in Equation 1 and β_3 is the estimate of average daily energy savings due to the program.

Equation 1: Post-Only Regression Model

$$ADC_{kt} = \sum_J \beta_{1j} Month_{jt} + \sum_J \beta_{2j} Month_{jt} \cdot ADClag_{kt} + \beta_3 Participant_k + \varepsilon_{kt}$$

Where:

ADC_{kt} = The average daily usage in kWh for customer *k* during billing cycle *t*. This is the dependent variable in the model

$Month_{jt}$ = A binary variable taking a value of 1 when *j*=*t* and 0 otherwise[26]

$ADClag_{kt}$ = Customer *k*'s energy use in the same calendar month of the pre-program year as the calendar month of *t*

[26] If there are T post-program months, there are T monthly dummy variables in the model, with the dummy variable $Month_{tt}$ the only one to take a value of 1 at time t. These are, in other words, monthly fixed effects.

$Participant_k$ = A binary variable indicating whether customer *k* is in the participant group (taking a value of 1) or in the control group (taking a value of 0)

ε_{kt} = The cluster-robust error term for customer *k* during billing cycle *t*. Cluster-robust errors account for heteroscedasticity and autocorrelation at the customer level

Variation-in-Adoption

The VIA method relies only on program participants to develop the counterfactual within the model,[27] which is made possible due to the program's rolling participation. Navigant will use the VIA analysis as a robustness check for the matching effort. A basic example regression specification of the VIA model is provided in Equation 2:

Equation 2: Baseline VIA Regression Model

$$ADU_{kt} = \alpha_k + \beta_t + \sum_{j=-\bar{m}}^{\bar{m}} \gamma_j D_{kt}^j + \varepsilon_{kt}$$

ADU_{kt} = Average daily or hourly energy use by household *k* in month *t*

α_i = Household-specific constant (fixed effect)

β_t = Month/year specific constant (fixed effect)

D_{kt}^j = A 0/1 indicator variable, taking a value of 1 if month *t* is the j^{th} month before/after household *k* installs the measure of interest. Month $\bar{m}=0$ is the month before enrollment

γ_j = Coefficient on the indicator variable D_{kt}^j

ε_{kt} = Model error term

Difference-in-Difference

A basic model specification for the DID method is provided in Equation 3, and α_3 is the estimate of average daily energy savings due to the program.

Equation 3: DID Regression Model

$$ADC_{kt} = \alpha_0 + \alpha_1 Post_t + \alpha_2 Participant_k + \alpha_3 Post_t \cdot Participant_k + \varepsilon_{kt}$$

Where:

[27] Harding, M. and A. Hsiaw. Goal Setting and Energy Conservation. July 2013. Available at: http://www.stanford.edu/~mch/resources/Harding_Goals.pdf.

ADC_{kt} = The average daily usage in kWh for customer *k* during billing cycle *t*. This is the dependent variable in the model

α_0 = Customer fixed effects

$Post_t$ = A binary variable indicating whether the calendar month *t* is before or after the measure install (taking a value of 1 after measure install and taking the value of 0 before measure install for each participant)

$Participant_k$ = A binary variable indicating whether customer *k* is in the participant group (taking a value of 1) or in the control group (taking a value of 0)

ε_{kt} = The cluster-robust error term for customer *k* during billing cycle *t*. Cluster-robust errors account for heteroscedasticity and autocorrelation at the customer level.[28]

[28] Ordinary Least Squares (OLS) regression models assume that the data are homoscedastic and not auto correlated. If either of these assumptions is violated, the resulting standard errors of the parameter estimates are incorrect (usually underestimated). A random variable is heteroscedastic when the variance is not constant. A random variable is auto correlated when the error term in one period is correlated with the error terms in at least some of the previous periods.

F.6.5 Comparison Group Limitations

Randomized experiments are the ideal method to control for non-program related changes in energy consumption and to specifically capture the savings from a given program, because they account for self-selection bias among program participants. However, randomized experiments require specific program designs up front and can be expensive to implement. When a randomized experiment is not possible, like in this study, researchers use a variety of approaches to simulate a randomized experiment (i.e., quasi-experimental), which is an active area of research. Regardless of the approach, a non-experimental design will be subject to self-selection bias.

Quasi-experimental approaches are often conducted using non-participants as a "reservoir" for selecting a comparison group. A comparison group is then selected as a subset from the reservoir of non-participants such that the selected non-participants are those that best represent the participants of interest (e.g. similar monthly energy consumption before program participation). There are various approaches for selecting a comparison group, which are all less preferable to a randomized experiment. Because these quasi-experimental approaches are imperfect, Navigant will choose one method for selecting a comparison group to serve in the final analysis of savings (in this case the post-only model with Euclidean matching), and Navigant will use additional methods for selecting a comparison group to serve as robustness checks on those results.

F.6.6 Proposed Data Filtering

In this section Navigant proposes a final set of data filters and arguments for the proposed filtering. However, and as mentioned earlier in this methods document, the methods may change as the project progresses.

Proposed Data Filtering

Filter	Proposed Filter for Primary Analysis	Argument
Non-Electric-Heat Sites	No filter	Sites with other heating fuels represent real world reductions in electric savings for BPA.
RMSE at some maximum threshold from VBDD fit	No filter	Removing sites based on data from their various heat signatures inherently removes sites who consume energy in an unexpected manner (e.g., less clear heat signature). Removing these sites leads to evaluation results that are more "prototypical" than real world. For example, efficiency measures may yield more comfort at the sacrifice of savings or some sites' consumption may be affected by "winterizing" vacation homes and Navigant intends to capture these unobservable effects in their evaluation results.
R^2 at some minimum threshold from VBDD fit	No filter	R^2 is higher for sites with larger slopes (R^2 is the proportional improvement compared to a flat line). As such, this filter would bias the population to sites with higher heating consumption and, presumably, higher savings.
Single vs Multi-Measure Installation Sites	No filter	Depreciating returns from multiple efficiency measures represent real world reductions in electric savings for BPA.
Sites w/ Net Metering	No filter	Installing efficiency measures at homes with on-site generation represents real world reductions in electric savings for BPA.
Sites w/ 0-kWh Bills	No filter	Based on utility feedback, bill periods with 0 usage typically represent vacant homes, most of which are due to "winterizing" vacation homes. This represents real world reductions in electric savings for BPA.
Filtered to Bills w/ <10x Median Usage	Yes filter	Sometimes meters are read wrong. This filter is expected to remove far fewer than 1% of bills.
Filtered to Bills Corresponding to Person & Place of Installation	Yes filter	Change of occupant can have an unexpected effect on consumption. Although the comparison group approach may account for this to some degree, Navigant supports this filter for the sake of simplicity.
Filtered to Sites w/ >=10 Months of Pre- and Post-Usage	Yes filter	Estimating saving as per HDD and per CDD enables Navigant to provide annual savings estimates without 12 months of data for every site. That said, too little data causes large error bands when extrapolating to the year.

Filter	Proposed Filter for Primary Analysis	Argument
Filtered to Sites w/ Quality Comparison Match	Yes filter	Without an experimental design, Navigant will use matched comparison sites to estimate the energy participants would have consumed without the program. In some cases, quality matches may be hard to find.
Filtered to Sites w/ Currently Available Application of Measure	Yes filter	With an emphasis on making forward looking improvements to the current UES, Navigant proposes to focus on sites that reflect currently eligible applications for the measures of interest.
Filtered to Sites w/ <50% Increase/ Decrease in Usage	No filter	Navigant proposes to use the comparison group to account for these drastic changes in energy consumption. Without more information on these sites, removing the outliers may lead to analyst bias, where results reflect the analyst's expectations rather than the actual data, especially given the arbitrary nature of setting such a threshold. Furthermore, these sites could receive further attention as a part of the secondary analysis.
Filtered to Sites w/o Missing Periods of Usage	No filter	Given the nature of the data collection process it is possible that bill periods were lost due to clerical error.

Appendix G: Changes to Measures over Time

Billing Analysis Methodology – Measure Changes over Time

To: BPA

Subject: Measure Changes over Time

Date: October 14, 2016

As a part of the 2016 impact evaluation of Bonneville Power Administration's (BPA) Unit Energy Savings (UES) portfolio, Navigant will further investigate savings from the data collected for the PTCS evaluation in 2013. This dataset includes PTCS duct sealing, commissioning controls and sizing and heat pump (HP) conversions. In this document Navigant outlines the major changes to these measures over time (as well as for ductless heat pumps, insulation and windows) and proposes approaches to address any significant changes.

G.1 Duct Sealing (DS)

Duct sealing has gone through several changes over the time period of received data. The bullets below contain the primary changes and a more detailed (although not exhaustive) list of changes can be found in Table G-1.

- **Specifications –** two versions of prescriptive and two versions of PTCS duct sealing specifications have been introduced over this time frame, although Navigant has primarily received data for which only one version of the prescriptive specification applies and one version of the PTCS specification applies. As such, no adjustments will be made to account for these changes and this will be considered a limitation of the received data for estimating future participant savings.
- **Application –** separate measures in the Implementation Manual were developed for duct sealing for new construction (single family or manufactured homes) and for low income manufactured homes. As such, Navigant will remove these applications when estimating savings for future participants.
- **QA/QC –** the extent of quality control on duct sealing installations appears to have changed over time. As these effects may have a lagging influence on savings (which would be hard to differentiate) Navigant proposes to compare savings for utilities with extensive and long-term QA/QC vs utilities with more limited QA/QC.

Table G-1: Duct Sealing Measure Changes over Time

Effective Date	Implementation Manual (IM) Change
Apr. 2010	Pre-installation test for DS no longer required for SF new construction
Oct. 2011	DS no longer labeled as "unbundled"
Apr. 2012	DS installation to comply with April 2009 version of "PTCS® Duct Technical Specifications"
Apr. 2012	DS measure for MH NC no longer included in IM
Apr. 2012	Duct systems sealed with aerosol sealant equipment according to the BPA-provided specifications will be considered PTCS-compliant
Apr. 2012	QA/QC on 10% of sites
Apr. 2013	No mention of QA/QC
Oct. 2013	IM allows2 DS measures per home for homes with 2 systems
Oct. 2014	Explicit mention of QA/QC again
Oct. 2014	Prescriptive DS described in IM
Oct. 2015	PTCS and Prescriptive DS measures for New Construction Single-Family homes expire
Apr. 2015	PTCS DS measures for manufactured homes (low income) will no longer be available
Apr. 2016	Prescriptive and PTCS to comply with DS specification dated April 1, 2015 rather than the 2014 version for Prescriptive and the 2009 version for PTCS
Apr. 2016	Pre–installation test requirements for DS no longer mentioned in IM

Source: Navigant

G.2 Commissioning Controls and Sizing (CCS)

CCS has gone through several changes over the time period of received data. The bullets below contain the primary changes and a more detailed (although not exhaustive) list of changes can be found in Table G-2.

- **Specifications –** three versions of CCS specifications have been introduced over this time frame, although Navigant has primarily received data for which only one version of the specification applies. As such, no adjustments will be made to account for these changes and this will be considered a limitation of the received data for estimating future participant savings.
- **Application –** CCS is no longer applicable for new construction in manufactured homes. As such, Navigant will remove these applications when estimating savings for future participants.
- **QA/QC –** the extent of quality control on CCS installations appears to have changed over time. As these effects may have a lagging influence on savings (which would be hard to differentiate) Navigant proposes to compare savings for utilities with extensive and long-term QA/QC vs utilities with more limited QA/QC.

Table G-2: CCS Measure Changes over Time

Effective Date	Implementation Manual (IM) Change
Apr. 2010	CCS for new construction MH expires
Apr. 2010	HP's must meet federal requirements for HSPF and SEER
Apr. 2011	No longer explicit mention that CCS can be installed in unlimited numbers per home
Oct. 2011	CCS no longer labeled explicitly as "unbundled"
Oct. 2011	Aux. lockout changed from 30 to 35F
Apr. 2012	CCS can no longer be bundled w/ HVAC measures other than DS
Apr. 2012	CCS installations to comply with ASHP installation specs 2007
Apr. 2012	Additional settings criteria are added to IM (new items are underlined) - The PTCS technician must correctly size the system to a 30-35 degree balance point, test for sufficient air flow across the coils and install an auxiliary heat lockout for when the outdoor temperature is above 35 degrees Fahrenheit. Exception: If the minimum setting on the thermostat is 40 degrees, 40 degrees may be used
Oct. 2013	CCS installations to comply with ASHP installation specs 2011
Oct. 2014	CCS installations to comply with ASHP installation specs 2013
Apr. 2015	CC relabeled as CCS
Apr. 2016	IM explicitly mentions QA/QC requirement
Apr. 2016	Explicit mention that HP can be replaced in a home "with or without air conditioning"
Apr. 2016	Explicit mention of limit to 2 per home (with certain size & system criteria)
Oct. 2016	Explicit mention that refrigerant charge is added to the list of specs (e.g., aux. lockout temp) although details are not given in the IM itself

Source: Navigant

G.3 Heat Pump (HP) Conversion

HP conversions have gone through several changes over the time period of received data. The bullets below contain the primary changes and a more detailed (although not exhaustive) list of changes can be found in Table G-3.

- **Specifications –** three versions of HP conversion specifications have been introduced over this time frame and minimum HSPF requirements were introduced. Unfortunately, however, Navigant has primarily received data for which only one version of the specification applies and this data was collected before the HSPF requirements were introduce. As such, no adjustments will be made to account for these changes and this will be considered a limitation of the received data for estimating future participant savings.
- **Application –** Historically duct sealing could be installed with HP conversions as two measures or as a bundled measure (e.g., "HP Conversion w/ DS"). Navigant proposes to address this change by treating "HP Conversion w/ DS" the same as a HP conversion where duct sealing is also performed.

- **QA/QC –** the extent of quality control on duct sealing installations appears to have changed over time. As these effects may have a lagging influence on savings (which would be hard to differentiate) Navigant proposes to compare savings for utilities with extensive and long-term QA/QC vs utilities with more limited QA/QC.

Table G-3: HP Conversion Measure Changes over Time

Effective Date	Implementation Manual (IM) Change
Oct. 2009	Explicitly mentions replacing eFAF (could have potentially replaced zonal before this date)
Apr. 2010	DS required if 50% (rather than 75%) of ducts are in unconditioned space, unless the ducts were previously certified or a PTCS duct leakage test indicates that the pre-existing duct leakage is too low to qualify for the PTCS duct sealing reimbursement
Apr. 2010	No longer explicit mention of CC requirement in conjunction w/ HP conversion
Oct. 2011	HP required at 8.5 HSPF to 9.0 HSPF depending on application
Apr. 2012	Aerosol sealant required on PTCS work
Apr. 2012	HP installations to comply with ASHP installation specs 2007
Oct. 2013	HP installations to comply with ASHP installation specs 2011
Apr. 2014	Electric hydronic baseboard allowed to count as eFAF for HP conv.
Apr. 2014 or Oct. 2014	HP conv. becomes a stand-alone measure w/o duct testing – i.e., eFAF can be converted to HP w/o duct testing, where it used to be required and contractors were required to seal ducts if tests indicated it was needed
Oct. 2014	HP installations to comply with ASHP installation specs 2013
Apr. 2015	PTCS HP measures with “ducts required” will no longer be available - e.g., HP conv w/ DS; but I *believe* HP conv. w/o DS can be done at a site where a separate DS measure is also done
Apr. 2016	Explicit mention of CC requirement in conjunction with HP conv
Apr. 2016	QA/QC mentioned as a requirement
Apr. 2016	Explicit mention that HP can replace a home "with or without air conditioning"

Source: Navigant

G.4 Ductless Heat Pumps (DHP) Replacing Electric Forced Air Furnaces (eFAF)

Ductless heat pump measures have gone through minor changes over the time period of received data. The bullets below contain the primary changes and a more detailed (although not exhaustive) list of changes can be found in Table G-4.

- **Application –** the implementation manual states that this measure is not applicable for new construction. As such, Navigant will remove these applications when estimating savings for future participants.

Table G-4: Changes for Ductless Heat Pumps Replacing Electric Forced Air Furnaces

Effective Date	Implementation Manual (IM) Change
Apr. 2014	DHP added as a low-income measure
Oct. 2014	DHP meet a HSPF requirement of 9.0 for single head systems and 8.0 for multi-head systems
Apr. 2015	IM no longer states that DS with DHP must meet min leakage requirements and must report if DS was done as '"DS-ASHP"
Apr. 2016	IM explicitly mentions for existing SF/MH only - no NC
Apr. 2016	IM no longer explicitly mentions requirement for contractor to receive training from manufacturer
Apr. 2016	IM no longer explicitly mentions limit for DHP of 1 per home

Source: Navigant

G.5 Insulation

Insulation measures have gone through minor changes over the time period of received data. The bullets below contain the primary changes and a more detailed (although not exhaustive) list of changes can be found in Table G-5.

- **Specifications –** two versions of the insulation installation specifications have been introduced over this time frame, and Navigant has received data for participants with installations using the 2014 specifications and the prior specifications. As such, Navigant will compare savings for these two groups to decide whether any adjustments are needed to reflect future participant savings.
- **Application –** the implementation manual states that this measure is not applicable for new construction. As such, Navigant will remove these applications when estimating savings for future participants.

Table G-5: Insulation Measure Changes over Time

Effective Date	Implementation Manual (IM) Change
Oct. 2013	Insulation not offered in all heat zones
Oct. 2014	Requires insulation to be installed to 2014 BPA Residential Weatherization Specifications
Oct. 2014	Open cavity or unfinished framed walls (e.g. knee walls in attics) must fill to a minimum of R-13 rather than R-15
Apr. 2016	IM provides flexibility with final R-value reported - "final R-value for reportable measure or max possible"
Oct. 2016	Requires insulation to be installed to 2016 BPA Residential Weatherization Specifications

Source: Navigant

G.6 Windows

Windows measures have gone through minor changes over the time period of received data. The bullets below contain the primary changes and a more detailed (although not exhaustive) list of changes can be found in Table G-6.

- **Specifications –** two versions of the windows installation specifications have been introduced over this time frame, and Navigant has received data for participants with installations using the 2014 specifications and the prior specifications. As such, Navigant will compare savings for these two groups to decide whether any adjustments are needed to reflect future participant savings.
- **Application –** the implementation manual states that changes may have occurred to this measure when applied to multi-family. As multi-family is outside the scope of this evaluation, Navigant will make no adjustments to their analysis to reflect these IM changes.

Table G-6: Windows Measure Changes over Time

Effective Date	Implementation Manual (IM) Change
Oct. 2014	Requires insulation to be installed to 2014 BPA Residential Weatherization Specifications
Apr. 2016	IM no longer explicitly states "Multifamily does not qualify for 0.22 windows or 0.30 patio doors at this time"
Oct. 2016	Requires insulation to be installed to 2016 BPA Residential Weatherization Specifications

Source: Navigant