NAVIGANT

ComEd
Advanced Thermostat Evaluation
Research Report

FINAL

Presented to
Commonwealth Edison Company

November 2, 2018

Prepared by:

Pace Goodman
Navigant

Will Sierzchula
Navigant

Carly Olig
Navigant

Mack Shaughnessy
Navigant

www.navigant.com

ComEd Advanced Thermostat Evaluation Research Report

Submitted to:

ComEd
Three Lincoln Centre
Oakbrook Terrace, IL 60181

Submitted by:

Navigant Consulting, Inc.
150 N. Riverside, Suite 2100
Chicago, IL 60606

Contact:

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

Jeff Erickson, Director
608.497.2322
Jeff.Erickson@Navigant.com

Josh Arnold, Associate Director
608.497.2328
Josh.Arnold@Navigant.com

TABLE OF CONTENTS

1. Executive Summary 1
2. Introduction 1
3. Measure Description 2
4. Research Methodology 3
    4.1 Methods Development 3
    4.2 Analysis Approach 4
5. Advanced Thermostat Evaluation Research Results 5
    5.1 Savings Estimates 5
        5.1.1 IL TRM 5
        5.1.2 The Effect of HERs on Advanced Thermostat Savings 6
        5.1.3 Different Baseline Thermostats 7
    5.2 Robustness and Uncertainty of the Results 8
6. Evaluation Research Findings and Recommendations 9
Appendix A. Evaluation Research Timeline 10
Appendix B. Evaluation Research Methodology 11
    B.1 Stakeholder Influence on Methods 11
    B.2 Process Billing, Tracking and Weather Data 12
    B.3 Develop Comparison Group 14
    B.4 Conduct Regression Analysis 16
    B.5 Estimate Savings 17
Appendix C. Sensitivity Analyses 19
    C.1 Sensitivity to Different Matched Comparison Groups 19
    C.2 Sensitivity to Model Specification 26
    C.3 Sensitivity to Alternate Data Processing 27
Appendix D. Additional Materials Supporting Illinois TRM Updates 29
    D.1 Signs of Self-Selection Bias Prior to Program Participation 29
    D.2 Results for Non-HER Participants in Comparison Test Group 3 Using Various Models 31
    D.3 Analysis of Heating and Cooling Loads Between Comparison Groups 35
Appendix E. Stakeholder Comments 36
    E.1 Summary of Comments Received 36
    E.2 Summary of Navigant's Response to Comments 36
    E.3 Comments Received 41

LIST OF TABLES AND FIGURES

Figure 4-1. Evaluation Research Overview 4
Figure 4-2. Advanced Thermostat Evaluation Research Task Overview 4

Figure C-1. Matching Plot – PY8 Participant Group 21
Figure C-2. Sensitivity Analysis Matching Plot – Comparison Test Group 1 23
Figure C-3. Sensitivity Analysis Matching Plot – Comparison Test Group 2 24
Figure C-4. Sensitivity Analysis Matching Plot – Comparison Test Group 3 26
Figure D-1. Comparison Test Group 3, Non-HER Participant Matching Plot 32
Figure E-1. Third Party Evaluation Findings where Savings are Presented per Heating or Cooling Energy Use 37
Figure E-2. Third Party Evaluation Findings where Savings are Presented per Whole Home Energy Use 38

Table 3-1. IL TRM Heating and Cooling Percent Reductions for Advanced Thermostats 2
Table 3-2. Participant Overview.....3
Table 5-1. Advanced Thermostat Research Estimated and TRM Estimated Savings for Participants Not Receiving HERs.....6
Table 5-2 .Advanced Thermostat Research Estimated Savings by HER Status 7
Table 5-3. Advanced Thermostat Research Estimated Savings by Baseline Thermostat Mode.....8

Table A-1. Evaluation Research Timeline.....10
Table B-1. Site Attrition Due to Data Processing.....13
Table B-2. Calculating Annualized Average Treatment Effect.....17
Table C-1. Overview of Matching and Observations Requirements in Sensitivity Tests.....19
Table C-2. Sensitivity Analysis Results Using Different Matching and Observation Thresholds 20
Table C-3. Regression Output – PY8 Participant Group 21
Table C-4. Matching Plot Data – PY8 Participant Group.....22
Table C-5. Sensitivity Analysis Regression Output – Comparison Test Group 1.....22
Table C-6. Sensitivity Analysis Matching Plot Data – Comparison Test Group 1 23
Table C-7. Sensitivity Analysis Regression Output – Comparison Test Group 2.....24
Table C-8. Sensitivity Analysis Matching Plot Data – Comparison Test Group 2 25
Table C-9. Sensitivity Analysis Regression Output – Comparison Test Group 3.....25
Table C-10. Sensitivity Analysis Matching Plot Data - Comparison Test Group 3 26
Table C-11. Robustness Model Specifications 27
Table C-12. Sensitivity Analyses Using PY8 Participant Group 27
Table C-13. Sensitivity to Alternate Data Processing – Savings Estimates for the PY8 Participant Group with Weather Data Matched by Bill Date 28
Table C-14. Sensitivity to Alternate Data Processing – Regression Output for the PY8 Participant Group with Weather Data Matched by Bill Date 28
Table D-1. Time Period Comparison – PY8 Participant Group Regression Output 30
Table D-2. Time Period Comparison – Comparison Test Group 3 Regression Output.....30
Table D-3. Comparison Test Group 3, Non-HER Participant Estimated Savings Between Models 31
Table D-4. Comparison Test Group 3, Non-HER Participant Site Attrition.....32
Table D-5. Comparison Test Group 3, Non-HER Participant Matching Data.....33
Table D-6. Comparison Test Group 3, Non-HER Participant Regression Output – With Post*Treatment Term.....34
Table D-7. Comparison Test Group 3, Non-HER Participant Regression Output – Without Post*Treatment Term.....34
Table D-8. PY8 Participant Group Matches Weather Variable Regression Output.....35
Table D-9. Comparison Test Group 3 Matches Weather Variable Regression Output.....35
Table E-1. Summary of Stakeholder Comments 36

1. EXECUTIVE SUMMARY

This report presents the results of Navigant's impact evaluation research of advanced thermostats rebated by (ComEd in Program Year 8 (PY8).[1] Advanced thermostats are defined in section 5.3.16 of the Illinois Technical Reference Manual (IL TRM) version 6.0.[2] This study is intended to provide an estimate of annual electric cooling savings from advanced thermostats installed through energy efficiency programs in Illinois to support data-driven IL TRM updates. However, the IL TRM administrator and Technical Advisory Commission (TAC) are responsible for updating the IL TRM. Thus, this research is offered for consideration by those groups and does not represent explicit updates for the IL TRM.

This analysis focused on electric savings related to energy used for cooling and did not study other potential benefits of the technology such as leveraging advanced thermostats for customer engagement opportunities or the role of advanced thermostats in demand response programs. Navigant is currently conducting separate research in IL examining potential benefits of thermostat optimization programs[3] and general research around non-energy impacts.[4]

Navigant's evaluation research indicates that advanced thermostats rebated in PY8 in ComEd's service territory save:

- Less total annual electric energy than IL TRM v6.0 specifies
- About as much electric heating energy as IL TRM v6.0 specifies (primarily from furnace fans)
- Less electric cooling energy than IL TRM v6.0 specifies

As such, Navigant recommends that the VEIC and IL TRM TAC reference 2% cooling reduction as the finding from this study most applicable for informing any updates to the IL TRM.[5]

For future work, Navigant is preparing scope for a study that will leverage advanced metering infrastructure for advanced thermostat evaluation research using more recent participants. Navigant will coordinate with the Advanced Thermostat Subcommittee for that work.

Further detail on these findings and recommendations can be found in Section 6

2. INTRODUCTION

Navigant's evaluation research of advanced thermostats installed in the ComEd service territory during PY8 is a follow-up to Navigant's 2016 evaluation of ComEd's advanced thermostat pilot. In 2016, the technology was less mature and likely bought and installed by early adopters. Thus, the findings may have been unrepresentative of future program years. In this updated evaluation research, Navigant analyzed the savings achieved by advanced thermostats incentivized through the PY8 Home Energy Assessment Program and Heating, Cooling, and Weatherization Rebate Program. This document summarizes the advanced thermostat measure's energy impacts; the methodology Navigant used to

[1] June 1, 2015 to May 31, 2016
[2] http://ilsagfiles.org/SAG_files/Technical_Reference_Manual/Version_6/Final/IL-TRM_Effective_010118_v6.0_Vol_3_Res_020817_Final.pdf
[3] ComEd is currently running two thermostat optimization pilots with Nest and Whisker Labs.
[4] Navigant is currently conducting research for ComEd to update how non-energy impacts are quantified and included in cost effectiveness tests in IL.
[5] Navigant also shared other possible findings that could inform the IL TRM in its comments submitted to VEIC May 24, 2018.

arrive at those figures; and sensitivity test results, which provide an indication of the findings' robustness and the limitations of the study.

The primary objective of this research was to update the "cooling reduction factor", defined in the Illinois Technical Reference Manual (IL TRM), which is used to calculate advanced thermostat annual electric savings. Navigant also provides information related to electric heating savings. Navigant did not estimate demand savings as a part of this study, nor did it investigate other possible benefits of advanced thermostats.

During this analysis, the EPA started certifying advanced thermostats under ENERGY STAR®.[6] This study did not include a rigorous investigation of their certification method.

3. Measure Description

As the IL TRM describes, an advanced thermostat reflects the "replacement of a manual-only or programmable thermostat, with one that has the default enabled capability —or the capability to automatically—establish a schedule of temperature setpoints according to driving device inputs above and beyond basic time and temperature data of conventional programmable thermostats." [7]

The IL TRM specifies this measure with heating and cooling reduction values (in units of savings per heating or cooling load) and customer specific heating and cooling load estimates, which vary based on home type, IL climate zone, heating system, AC efficiency, and participant characteristics. At a basic level, the product of the heating or cooling reduction and the heating or cooling load is measure savings. The IL TRM currently specifies two heating reduction values, one for replacing manual thermostats and one for replacing programmable thermostats. However, based on research[7], the IL TRM only specifies one cooling reduction value for advanced thermostats replacing either manual or programmable thermostats. Table 3-1 presents the IL TRM's heating and cooling percent reductions.

Table 3-1. IL TRM Heating and Cooling Percent Reductions for Advanced Thermostats

Weather Savings	Baseline Thermostat	Percent Reduction Value
Heating	Programmable	5.6%
Heating	Manual	8.8%
Cooling	Programmable or Manual	8%

Source: IL TRM v6.0

In Navigant's analysis, the participant set comprised 10,105 accounts who received an advanced thermostat rebate in PY8, after data processing and matching. Table 3-2 provides an overview of those accounts based on factors including home type, and proportion of homes that replaced programmable thermostats. For more information on site attrition, see Section B.2.

[6] https://nest.com/blog/2017/02/28/the-nest-thermostat-earns-an-energy-star/
[7] http://ilsagfiles.org/SAG_files/Technical_Reference_Manual/Version_6/Final/IL-TRM_Effective_010118_v6.0_Vol_3_Res_020817_Final.pdf

Table 3-2. Participant Overview

Participant Category	Participant Population PY8*	Analysis Dataset*	Subset of Analysis Dataset Unaffected by HER*
Number of Participants	23,944	10,105	2,641
Natural Gas Heat	96%	96%	96%
Central AC	99%	99%	99%
Programmable Thermostat Baseline	80%	84%	81%
Programmed Programmable Thermostat Baseline	44%	51%	50%
Single Family Homes	90%	93%	89%
Existing, rather than New, HVAC Systems	93%	94%	94%
Received One Rebate	95%	95%	95%
Nest Thermostat	80%	84%	85%
Ecobee Thermostat	20%	16%	15%
Received Home Energy Reports	57%	74%	0%

*The percentages represent the proportion of the dataset where the participant category could be consistently assigned based on program tracking data.
Source: ComEd billing and tracking data and Navigant team analysis.

4. Research Methodology

Navigant conducted this research by first developing an analytical method through stakeholder engagement, and then employing that approach to provide results the IL TRM administrator could use to support data-driven updates to the IL TRM.

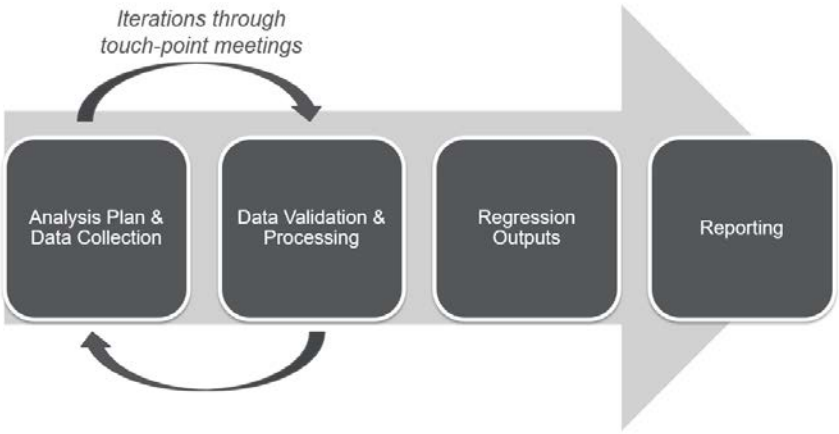
4.1 Methods Development

Navigant's preferred research approach is to develop an agreed-upon methodology before calculating results. Navigant considers this two-step process to be best practice in research, aligning with scientific principles. Additionally, this approach is intended to reduce the effects of confirmation bias and make work transparent and repeatable. Figure A-1 presents a flow chart outlining the analysis' phases.

Navigant designed the advanced thermostat study to reach agreement on the methods prior to performing the analysis and providing results. This study involved coordination and communication with the Advanced Thermostat Subcommittee, a subcommittee within the IL TRM Technical Advisory Commission (TAC). Coordination primarily occurred throughout the planning, data collection, and methods development phases. Feedback not incorporated into the study was either identified for consideration in future research, or Navigant provided a response articulating our position on the issue. Please see Table A-1 for a detailed timeline of the evaluation research. See Section B.1 for examples of stakeholder feedback that Navigant incorporated into the study. The IL TRM administrator is responsible

for coordinating updates to the IL TRM in response to this research and Navigant will continue to include the IL TRM TAC Advanced Thermostat Subcommittee in any relevant future research.

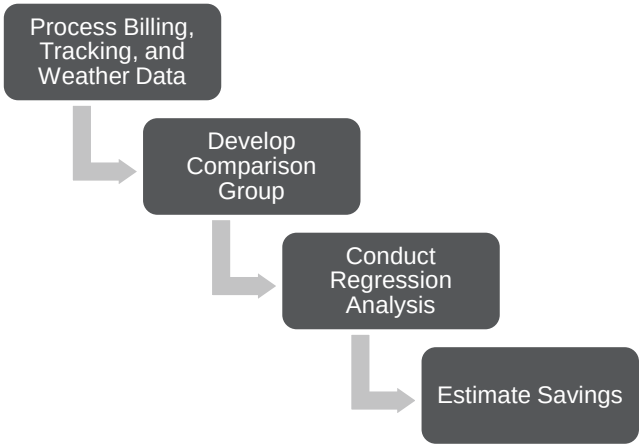
Figure A-1. Evaluation Research Overview



4.2 Analysis Approach

The methodology for this study included four primary tasks, shown in Figure A-1. Further details on each task can be found in Appendix B.

Figure A-1. Advanced Thermostat Evaluation Research Task Overview



Navigant collected, processed, and validated the following datasets – customer billing, program tracking, and weather. Then, Navigant combined these datasets for analysis. Further details on processing tasks for each dataset can be found in Section 6.B.2.

Navigant developed the comparison group by finding, for each participant, the control account[8] with the most similar energy use during the matching period (June 2013 – May 2014), within the same zip code and Home Energy Report (HER) wave.[9] More details can be found in Section B.2.

Navigant estimated advanced thermostat savings using a Linear Fixed Effects Regression (LFER) model. This model compares energy use before and after treatment for customers who did and did not receive advanced thermostat rebates. This type of model was used in an Xcel Smart Thermostat evaluation, to evaluate smart thermostat savings in Michigan, and was recommended by subcommittee members during the methods development.[10] [11] [12] Section B.3 provides an overview of comparison group selection and a timeline of the pre-treatment, matching, and post-treatment periods; Sections B.4 and 6.B.5 provide further detail on the savings estimation methodology.

5. Advanced Thermostat Evaluation Research Results

Navigant provides results separately for (1) savings estimates, and (2) analysis outputs indicating the robustness and uncertainty of the results.

5.1 Savings Estimates

This study yielded results relevant to the IL TRM, the effect of HERs on advanced thermostat savings, and variation in savings for different baselines (e.g., programmable and manual thermostats).

5.1.1 *IL TRM*

For informing updates to the IL TRM, Table 5-1 compares savings estimated by Navigant's evaluation research for customers who did not receive HERs (see Section 5.1.2 for details) with savings calculated using the IL TRM v6.0. Navigant's evaluation research of advanced thermostats for PY8 participants in ComEd's service territory yielded a cooling reduction value of 2%. Because this type of analysis estimates net savings, no further net-to-gross (NTG) adjustment is necessary.[13] This analysis estimates cooling savings that are not statistically different than zero but are statistically different than the 8% defined in IL TRM v6.0 at the 90% confidence level. In addition to the cooling value, this study estimated a heating reduction value of 5.8%. The heating savings estimate is statistically different than zero and is in line with the values in the IL TRM. The total annual electric savings estimate (110 kWh per site per year) is statistically different than zero, and statistically different than the IL TRM's specified total savings for the PY8 participants.

[8] Control accounts did not receive an advanced thermostat incented by ComEd but may have installed an advanced thermostat on their own without obtaining a utility rebate. As such, Navigant considers the savings estimates from this study to be net savings.
[9] An HER wave refers to a group of customers who were enrolled in the HER program at the same time.
[10] Xcel Energy Evaluation: https://www.xcelenergy.com/staticfiles/xe-responsive/Company/Rates%20&%20Regulations/Regulatory%20Filings/CO-Smart-Thermostat-Pilot-Evaluation.PDF
[11] MI Evaluation Research: https://www.michigan.gov/documents/mpsc/Tier_3_Tstat_Calibration_Study_EWR_Presentation_623038_7.pdf
[12] Subcommittee members stated in their comments December 13, 2016, "For evaluating cooling and heating savings, the model should account for how weather affects energy use and how the thermostat affects this relationship. The CITS model accounts for these factors directly whereas the PO model does not, making CITS better suited for this analysis." Navigant's understanding is that the LFER model is comparable to CITS, which each differ from Navigant's preferred model – the post-only (PO) or lagged dependent variable model.
[13] The Illinois NTG Working Group is reviewing the relationship between billing analyses and net savings and Navigant will use those protocols for future studies once finalized.

Table 5-1. Advanced Thermostat Research Estimated and TRM Estimated Savings for Participants Not Receiving HERs

Savings Methodology	Sample Size	Cooling		Heating*		Total
		Percent Reduction (90% CI)	Electric Energy Savings by kWh/year/site	Percent Reduction (90% CI)	Electric Energy Savings by kWh/year/site	Electric Energy Savings by kWh/year/site (90% CI)
IL TRM v6.0	2,641	8%	170	6.2%†	96	266
Evaluation Research Findings	2,641	1.8% (-1.8 to 5.4%)	35	5.8% (1.5 to 10.2%)	74	110 (6 to 213)

*This evaluation focuses on electric savings, and the heating reduction corresponds in large part to furnace fan savings.
† Reflects a weighted average between the two IL TRM v6.0 heating reduction values for advanced thermostats replacing manual and programmable thermostats.
Source: ComEd billing and tracking data and Navigant team analysis.

5.1.2 *The Effect of HERs on Advanced Thermostat Savings*

ComEd's HER program affected 57% of ComEd residential advanced thermostat recipients in PY8. Additionally, three quarters of the customers included in the evaluation research data set after processing had received HERs.

The IL TRM should specify measure savings that are not diminished by HER, because HER program evaluations have a standard practice of removing savings attributed to other measures and programs, including advanced thermostats, through an "Uplift" analysis when reporting savings. This approach ensures that IL energy efficiency programs aren't double penalized for savings that overlap, and that savings aren't double counted. Navigant found 2% cooling savings for PY8 advanced thermostat participants who did not receive HERs and about 0, or even perhaps slightly negative cooling savings for participants who did receive HERs (see Table 5-2). Neither cooling savings estimate is statistically different than zero with 90% confidence, but each is statistically different than the IL TRM v6.0, which specifies 8% cooling reduction. Additionally, the cooling savings between HER and non-HER groups aren't statistically significantly different with 90% confidence, but the trend is consistent across sensitivity testing and is statistically significantly different with 70% confidence.

Table 5-2. Advanced Thermostat Research Estimated Savings by HER Status

Participant Group	Sample Size	Cooling		Heating*		Total
		Percent Reduction (90% CI)	Electric Energy Savings by kWh/year /site	Percent Reduction (90% CI)	Electric Energy Savings by kWh/year /site	Electric Energy Savings by kWh/year /site
All PY8 Advanced Thermostat Participants with Adequate Data	10,105	-0.3%	-7	6.4%	88	81
PY8 Participants Receiving Home Energy Reports (HERs)	7,464	-0.9% (-2.8 to 0.9%)	-21	6.5% (4.3 to 8.7%)	93	72
PY8 Participants Not Receiving Home Energy Reports (HERs) – useful to Inform the IL TRM	2,641	1.8% (-1.8 to 5.4%)	35	5.8% (1.5 to 10.2%)	74	110

*This evaluation focuses on electric savings, and the heating reduction corresponds in large part to furnace fan savings.
Source: ComEd billing and tracking data and Navigant team analysis.

The participant group labeled as "All PY8 Advanced Thermostat Participants with Adequate Data" in Table 5-2 will be referred to as the "PY8 participant group," hereafter. The "PY8 Participants Not Receiving Home Energy Reports (HERs) – useful to Inform the IL TRM" participant group is a subset of the PY8 participant group and will be referred to as the "TRM-recommended participant group (excludes HER effects)," hereafter. The TRM-recommended participant group (excludes HER effects) was formed by filtering the "PY8 participant group" to only customers that did not participate in the HER program.

5.1.3 *Different Baseline Thermostats*

Navigant compared savings from advanced thermostats replacing different baseline thermostats. The IL TRM defines a single value for cooling reduction (8%) for advanced thermostats replacing manual or programmable thermostats. However, the TRM defines two heating savings values, one for a manual thermostat baseline (8.8%) and one for a programmable thermostat baseline (5.6%) replaced by an advanced thermostat. Navigant found consistent cooling savings regardless of whether participants replaced manual or programmed programmable thermostats but found much higher heating savings for participants who replaced manual thermostats compared to programmed programmable thermostats (see Table 5-3). These results have high uncertainty and variance, but don't indicate a need at this time for revising the IL TRM's framework of specifying two heating reduction values and only one cooling reduction value. This topic could be revisited in future work.

Table 5-3. Advanced Thermostat Research Estimated Savings by Baseline Thermostat

Participant Group	Sample Size	Cooling		Heating*	Total	
		Percent Reduction	Electric Energy Savings by kWh/year/site	Percent Reduction	Electric Energy Savings by kWh/year/site	Electric Energy Savings by kWh/year/site (90% CI)
All Sites*	10,105	-0.3%	-7	6.4%	88	81 (29 to 134)
Manual Thermostat Baseline	1,550	-0.7%	-16	10.7%	155	139 (7 to 271)
Programmed Programmable Thermostat Baseline	4,883	-0.5%	-11	4.9%	68	58 (-20 to 135)

*The results for manual baseline thermostats and programmed programmable thermostat baselines do not average to "all sites," because there are additional groups for which we have not isolated savings (e.g., programmable thermostats on hold and where this data is unknown).

Source: ComEd billing and tracking data and Navigant team analysis.

5.2 Robustness and Uncertainty of the Results

In this subsection, Navigant speaks to the limitations of this study and to our best effort to provide additional analysis outputs that can inform our interpretation of the findings given these limitations.

1. **Statistical Significance:** Navigant presents some savings estimates that are not significantly different than zero but are significantly different than the values in IL TRM v6.0 at the 90% confidence level. This level of uncertainty is not ideal but does provide an indication that updates to the IL TRM or future research is warranted.
2. **Self-Selection Bias:** Self-selection bias affects all regression-based approaches to estimating savings that do not include a randomized or experimental study design. This study did not employ a randomized design but used industry best-practices for non-experimental design approaches. In order to determine if self-selection was an issue, Navigant ran sensitivity analyses comparing participant and control usage for the two years prior to the treatment period and found no signs that the two groups were diverging. This result indicates self-selection bias did not strongly affect energy use prior to program participation (see Appendix D for more details). Furthermore, Navigant also developed multiple comparison groups, which provide an indication as to the sensitivity of the results to a single comparison group; Navigant found that these groups yielded consistent savings estimates. Navigant presents more details on the sensitivity analysis in Section C.1. While these additional analysis outputs don't prove the absence of self-selection, they do serve as additional rigor beyond standard practice, and show no signs of self-selection bias. These results indicate that if self-selection bias is affecting results, the effect would have to be equal across the various comparison groups and the effect would have to start occurring simultaneously with the program, having not been occurring prior.
3. **Past and Future Savings:** Evaluation research typically provides accurate estimates of savings retrospectively, which can inform energy efficiency program planning, but are imperfect predictors of the future. This challenge is not unique to advanced thermostats. This study is intended to provide an independent estimate of PY8 advanced thermostat electric savings to support data-driven IL TRM updates. The IL TRM administrator and the IL TRM TAC are responsible for IL TRM updates, and can use this information to make informed updates.

4. **Model Specification:** Model specification bias refers to possible erroneous results due to model misspecification. Matching to a comparison group is designed to limit model specification bias by making the comparison group as similar to the participants as possible based on observable characteristics. To investigate remaining model specification bias after matching in this study, Navigant tested seven total model specifications, which all yielded consistent findings. Navigant presents these sensitivity results in SectionAppendix C. These findings indicate that any possible imperfections in the model specification appear to have a limited effect on the savings estimate.

6. Evaluation Research Findings and Recommendations

The following describes key evaluation research findings and recommendations.

Finding 1. Electric cooling and total annual electric energy savings attributed to advanced thermostats are statistically different than what is defined in the IL TRM v6.0 at the 90% confidence level. Electric heating savings are not statistically different from the value in the IL TRM at the 90% confidence level.

Recommendation 1. Navigant recommends the IL TRM administrator and the IL TRM TAC consider updating the cooling reduction factor in the IL TRM v7.0. They can reference 2% cooling reduction as the finding from this study most applicable for informing any updates to the IL TRM.

Finding 2. Navigant found an indication that advanced thermostat participants who had been receiving Home Energy Reports saved less cooling energy than advanced thermostat participants who had not been receiving the reports, although this trend was not statistically significant at the 90% confidence level.

Finding 3. While not statistically significant with 90% confidence, homes that replaced manual thermostats with advanced thermostats tended to have equivalent cooling savings, but higher heating savings, compared to homes that replaced programmed programmable thermostats with advanced thermostats. This finding aligns with the IL TRM v6.0, which specifies two heating reduction values (one for manual thermostat baselines and one for programmable thermostat baselines), but only one cooling reduction value. These results have high uncertainty and variance, but don't indicate a need for revising the IL TRM's framework of specifying two heating reduction values and only one cooling reduction value. This topic could be revisited in future work.

APPENDIX A. EVALUATION RESEARCH TIMELINE

As shown by the dates in Table A-1. Evaluation Research Timeline

, this project progressed from initial proposed methods to publicly available regression results over about 22 months. During this time, the Advanced Thermostat Subcommittee used seven months to reach consensus on methods, data collection required 10 months due to the magnitude of the data, and Navigant used five months to conduct the analysis and deliver and discuss regression outputs with the subcommittee. For future projects involving engaged stakeholders and complex analysis, evaluators may consider more expeditious processes, such as requesting data before reaching agreement on methods. This approach would enable discussions on methods and data collection to happen in parallel but would limit stakeholders' ability to comment on the data request.

Table A-1. Evaluation Research Timeline

Date	Evaluation Event
June 17, 2016	Navigant shared the first draft of our detailed research plan with stakeholders.
By January 20, 2017	The subcommittee finalized the high-level evaluation methodology and data needs during a touch-point meeting and through e-mail.
February 2, 2017	Navigant requested energy use data from ComEd.
March 4, 2017	VEIC shared with the subcommittee that Navigant was not able to obtain thermostat data from manufacturers due to substantial and unanticipated challenges.
November 7, 2017	The subcommittee finalized the methodology for calculating usage reduction percentage (% reduction) from the regression outputs during a touch-point meeting.
December 18, 2017	Navigant received final evaluation data from ComEd.
December 19, 2017	Navigant validated evaluation data and confirmed data sufficiency.
February 7, 2018	The subcommittee validated comparison group matches through e-mail.
April 25, 2018	Navigant delivered regression outputs.
May 7 to June 19, 2018	The subcommittee discussed regression outputs, future research tasks and implication to the IL TRM version 7 during two touch-point meetings and a comment-and-response period.
August 27, 2018	Navigant delivered the draft report

Source: ComEd billing and tracking data and Navigant team analysis.

APPENDIX B. EVALUATION RESEARCH METHODOLOGY

This appendix provides detail on each of the primary tasks completed during the evaluation research as well as stakeholders' influence on methods.

B.1 Stakeholder Influence on Methods

This appendix subsection provides detail on the stakeholder engagement process for this study. Due to stakeholder feedback, Navigant adjusted the proposed methods and conducted additional robustness and sensitivity tests (see bulleted lists below). However, some comments, such as suggestions to use AMI data, could not be incorporated due to various constraints. For example, AMI was only available for 18% of ComEd's total meters in 2014, which would have affected evaluation of PY8 participants. Conducting this analysis with AMI would have reduced the sample size to a point where the new data stream would add uncertainty rather than reduce it.[14] Examples of Navigant's adjustments to the methods include the following:

- **Model Type:** Navigant compromised with one thermostat manufacturer to use a model specification aligned with their preferences. Navigant used a linear fixed effect regression (LFER) model in this analysis. Navigant initially proposed a post-only, or lagged dependent variable, model, where the manufacturer expressed preference for a LFER or comparative interrupted time series (CITS) model, as exemplified in the following quote: "For evaluating cooling and heating savings, the model should account for how weather affects energy use and how the thermostat affects this relationship. The CITS model accounts for these factors directly whereas the PO model does not, making CITS better suited for this analysis." – A Thermostat Manufacturer, December 13, 2016.
- **Specific Model Parameters:** Navigant compromised with a thermostat manufacturer on specific model parameters. Navigant originally proposed a model that produced savings only as those correlated with heating (i.e., post*treat*HDD) and those correlated with cooling (i.e., post*treat*CDD), but the thermostat manufacturer requested a "main treatment effect" (i.e., post*treat), which Navigant was willing to include in the model.[15]
- **Matching by Zip:** Navigant compromised with a thermostat manufacturer and the Environmental Law and Policy Center (ELPC) to restrict the comparison group matching to be within zip code. Navigant feels that if matching by zip code has a strong impact on energy use, matching on pre-installation energy use alone will inherently account for any benefits from this additional restriction. For example, ELPC stated in their comments December 13, 2016 "ELPC believes matches should be made with a higher level of geographic granularity, ideally matching within the same zip code."

[14] Navigant proposed to use AMI as of November 18, 2016, but agreed with a manufacturer's comments from December 13, 2016, which stated "We support using AMI data rather than monthly billing data but are also concerned with sample attrition and representativeness. What fraction of rebate participants will have a full year of pre-installation daily AMI data. If there is large attrition, then we have concerns about potential bias." Using AMI to evaluate PY8 participants could have reduced the sample size by almost 80% and the subcommittee agreed to use billing data at the touch-point meeting January 20, 2017.

[15] A manufacturer commented on December 13, 2016 "the proposed regression model (Equation 1) estimates savings only through an interaction term with cooling degree days (CDD70) and does not include any main treatment effect" and the subcommittee agreed on January 20, 2017 during the touch-point meeting to include this additional variable in the model.

- **Balance Temperatures:** At the request of the ICC staff, Navigant changed the balance temperatures used to calculate heating and cooling degree-days to reflect the IL TRM (60° for heating and 65° for cooling).

After presenting initial results, Navigant also conducted additional analysis and provided additional content as requested by the subcommittee. Examples of these additional items are included in the following:

- **ICC Staff Requests:** Navigant conducted analyses requested by the ICC staff on May 9, 2018. These analyses included additional investigation for signs of possible self-selection bias as well as unique estimates of savings for subsets of the population. The results of these analyses are presented in Navigant's second addendum shared May 22, 2018, which is also available on the TRM SharePoint.[16] These results are also included in Appendix D.
- **Thermostat Manufacturer Requests:** As requested by one thermostat manufacturer, Navigant provided monthly energy use and weather data to the subcommittee on May 24, 2018.

B.2 Process Billing, Tracking and Weather Data

The following subsections outline the steps taken to process each data source.

B.2.1 Billing Data

Prior to billing data processing, the dataset included 23,884 participants and 1,946,260 controls. The billing data was then processed by developing a dataset with unique observations by read date and account number. Billing data processing primarily involved removing duplicate records and accounts with insufficient data. Navigant also confirmed that the processed data aligned with data used in other evaluations where overlap existed.

[16] Navigant's presented slides and subsequent addendum can be found at the following link and are titled "ComEd - Adv Therm - Regr Outputs - DRAFT - 2018-04_25," "ComEd - Adv Therm - Regr Outputs Addendum - DRAFT - 2018-05-08," and "ComEd - Adv Therm - Regr Outputs Addendum2 - DRAFT - 2018-05-22" - https://portal.veic.org/projects/illinoistrm/Shared%20Documents/Forms/AllItems.aspx?RootFolder=%2fprojects%2fillinoistrm%2fShared%20Documents%2fWorking%20Group%20Materials%2fAdvanced%20Thermostat%20Subcommittee%2fComments%20to%20Navigant%27s%20Report%20and%20or%20Next%20Steps%20for%20TRM%20Update%20May%2029%202018&FolderCTID=0x01200042B0ABF3AA22EE4888A0EDE62AB5CED4

Table B-1. Site Attrition Due to Data Processing

Data Processing Step	Participants	Controls
Raw Data	23,884	1,946,260
Bad Reads (e.g., All Estimated Bills)	23,884	1,946,259
Averaged Bills in the Same Month	23,884	1,946,259
Removed Outliers (i.e., observations one order of magnitude above or below median usage)	23,884	1,946,234
Had 10+ Observations in Pre- and Post-Periods	13,303	1,331,410
Received No Other Rebates or Rebated Measures from Res HVAC or HEA Programs	11,571	1,281,579
Had Matches in same HER Wave and Zip	10,554	1,281,579
Had 10+ Observations in the Matching Period	10,105	1,239,002
Match Participants and Controls to Form Analysis Dataset	10,105	9,407

Source: ComEd billing and tracking data and Navigant team analysis.

B.2.2 Tracking Data

Tracking data from several program years of the Home Energy Assessment and Heating, Cooling and Weatherization Rebate programs was used to identify customers that received rebates for advanced thermostats as well as for other measures. Navigant removed accounts from the combined dataset (i.e., for participants and the comparison group) that received rebates for other measures, and for advanced thermostats in years other than PY8, so that measure-related energy impacts could be attributed solely to PY8 advanced thermostats.

B.2.3 Weather Data

To provide savings estimates for both an actual and a typical meteorological year, Navigant used actual weather data from the National Oceanic and Atmospheric Administration (NOAA) and typical weather data from Typical Meteorological Year (TMY3) datasets. The results presented in Section 5 reflect typical year savings. Data processing for weather data consisted of developing heating and cooling degree-days, which were used as terms in the regression model as well as to estimate annual savings. These weather data sources and associated processing steps are described in further detail in the following subsections.

Actual Weather Data
Navigant used actual weather (drybulb) values in the advanced thermostat regression model. Weather values came from NOAA quality controlled local climatological data weather stations. We used weather data from the closest weather station relative to the account's zip code centroid.[17, 18]

Navigant calculated Heating and Cooling Degree Days (HDD and CDD) hourly with balance temperatures of 60 and 65 respectively. For example, if a weather station had an hourly temperature of 55, it would have a Heating Degree Hour (HDH) value of 5 for that hour. These hourly HDH and CDH values were then summed up for each day, and divided by 24 to get daily HDD and CDD values. Navigant then summed up daily HDD and CDD values over each month, and divided that figure by the number of days in that month to approximate average monthly HDD and CDD per day. We combined weather and usage

[17] Because participants and controls were matched on zip code, both matched accounts had the same weather data for the analysis.
[18] In instances where a station was missing data, we filled those values with weather from O'Hare International Airport.

data by year and calendar month. Navigant also combined these data by read date in some of its sensitivity testing of the results.

Typical Meteorological Year Data
Navigant used TMY3 data to estimate program savings for a typical weather year. This data provides hourly meteorological values that typify weather for a specific location and calendar month.[19] TMY3 values allow Navigant to estimate energy savings in absolute values for typical weather years. Navigant calculated average annual HDD and CDD values based on monthly figures using the method described in the previous subsection. Navigant multiplied TMY HDD and CDD values by the treatment effect coefficients, which incorporated weather values, as described in Section B.4.

B.3 Develop Comparison Group

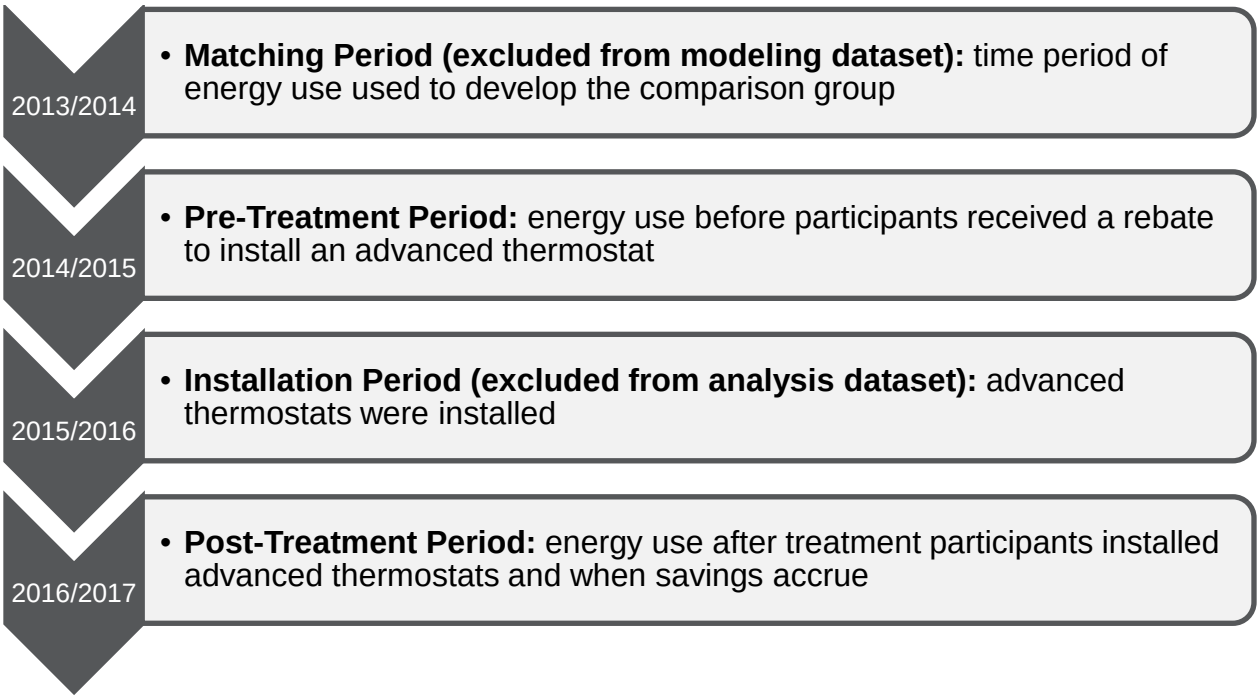
Navigant developed a comparison group by matching participant usage (customers with advanced thermostats) to a pool of potential controls (customers without advanced thermostat rebates) based on similar energy usage. The pool of non-participant households available for matching consisted of approximately two million ComEd residential customers.[20] Navigant matched participants to controls within the same zip code and HER wave to control for additional factors which could influence energy savings. The purpose of this matching method was to develop a comparison group that best reflects the participant group based on the differences in energy use between a participant and control in the period before the participant upgraded their thermostat.[21] Navigant refers to the period before the participant upgraded their thermostat, when energy consumption data was used to develop the comparison group, as the "matching period." Figure B-1 provides an overview of the four discrete time periods relevant to the analysis.

[19] Wilcox, S and W. Marion, 2008. *Users Manual for TMY3 Data Sets*. NREL/TP581-43156.

[20] This number includes accounts that did and did not receive HER.

[21] The quality of a match is denoted by the Euclidean distance to the participant. This distance is measured as the mean squared of the difference in monthly energy use between a participant and a potential match in terms of pre-usage over the matching period. The non-participant customer with the shortest Euclidean distance to a participant is chosen as the matched comparison for the participant. Matching, for this study, was done with replacement and the standard error accounted for this by using a robust standard error that clustered the error around the individual at every instance of each individual.

Figure B-1. Advanced Thermostat Evaluation Research Time Periods



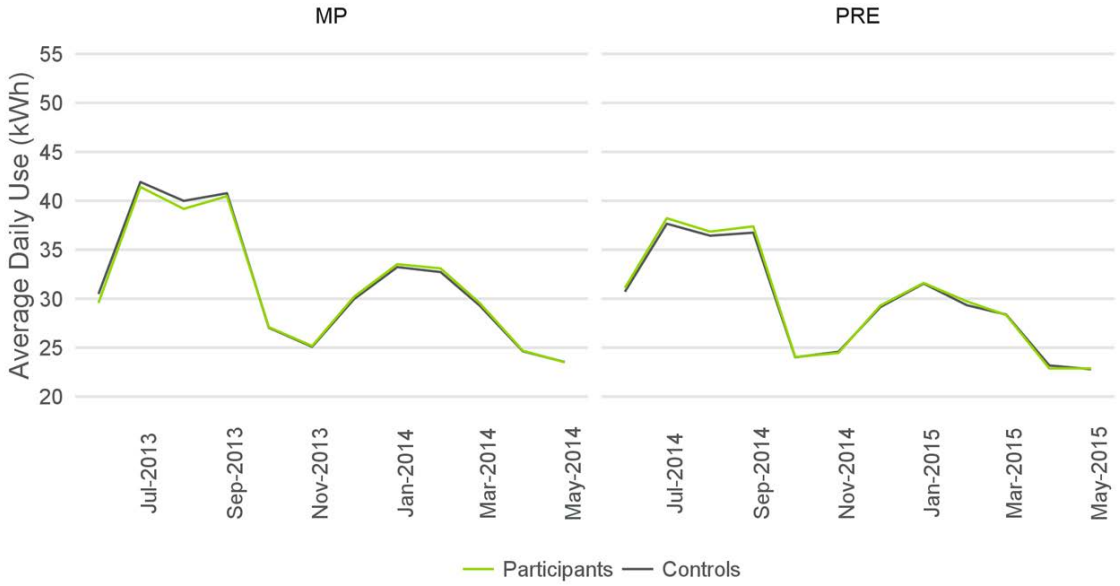
Navigant validated the matched comparison group by visualizing average monthly usage for participants and controls during the matching period. Navigant compared usage during the pre-treatment period as a secondary test of goodness-of-fit. Figure B-2 shows the of average monthly usage for participants and controls in the matching and pre-treatment periods for the PY8 participant group. Since the matching method did not include usage during pre-treatment months (June 2014 – May 2015), similar participant and control usage in that period would suggest the matches are performing well. Navigant's selection of matches used for regression was supported by participant and control usage being approximately the same during *both* the matching *and* pre-treatment periods. Furthermore, this approach is consistent with the academic literature.[22,23,24] Navigant also conducted additional analysis to determine that groups used energy similarly before advanced thermostats were rebated and found similar savings estimates across multiple comparison groups. Plots of the matched comparison groups used in these sensitivity analyses are in Section C.1.

[22] Stuart, E.A. and Rubin, D.B. 2007. Best Practices in Quasi-Experimental Designs: Matching methods for causal inference. Chapter 11 (pp. 155-176) in *Best Practices in Quantitative Social Science*. J. Osborne (Ed.). Thousand Oaks, CA: Sage Publications. (GS citations: 131)

[23] Ho, D.E., K. Imai, G. King, and E.Stuart. 2007. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis* 15(3): 199-236. (GS citations: 1590). Winner of Warren Miller Prize for best paper published in *Political Analysis* in 2007

[24] Imbens, G.M. and J.M Wooldridge.2008. "Recent developments in the econometrics of program evaluation". National Bureau of Economic Research, working paper no. 14251. http://www.nber.org/papers/w14251. (GS citations: 2029).

Figure B-2. PY8 Participant Group Matching Plot



Source: ComEd billing and tracking data and Navigant team analysis.

B.4 Conduct Regression Analysis

To estimate energy savings, Navigant used a Linear Fixed Effects Regression (LFER) model. In this model, average daily kWh consumption by household *k* in bill period *t* is denoted by $ADU_{k,t}$. Formally, the expression of the LFER model is shown in Equation 1.

Equation 1. Linear Fixed Effects Regression Model

$$ADU_{k,t} = \alpha_k + \beta_1 HDD_{60,k,t} + \beta_2 CDD_{65,k,t} + \beta_3 post_t + \beta_4 post_t * HDD_{60,k,t} + \beta_5 post_t * CDD_{65,k,t} + \beta_6 treat_k * HDD_{60,k,t} + \beta_7 treat_k * CDD_{65,k,t} + \beta_8 post_t * treat_k + \beta_9 post_t * treat_k * HDD_{60,k,t} + \beta_{10} post_t * treat_k * CDD_{65,k,t} + \varepsilon_{k,t}$$

Where:

α_k = Site fixed effects, which are binary variables (one for each site) that take on the value of 1 for a given site, *k,* and 0 otherwise. This variable accounts for site specific conditions that do not change over time, such as the number of occupants.

$HDD_{60,k,t}$ = Heating degree days for customer k during time (i.e., bill) *t* at a 60°F balance temperature.

$CDD_{65,k,t}$ = Cooling degree days for customer k during time (i.e., bill) *t* at a 65°F balance temperature.

$post_t$ = A binary variable indicating whether time period *t* is after the advanced thermostat installation (taking a value of 1) or before (taking a value of 0). This variable will take values of 1 and 0 for both participant and comparison group sites.

$treat_k$ = A binary variable indicating whether customer *k* is in the treated participant group (taking a value of 1) or in the comparison group (taking a value of 0). This variable will not change over time for any customers.

$\varepsilon_{k,t}$ = The cluster-robust error term for customer *k* during time period *t*. Cluster-robust errors account for heteroscedasticity and autocorrelation at the customer level.

Three observations about this specification deserve comment. First, the coefficient α_k captures all household-specific effects on energy use that do not change over time, including those that are unobservable. The effect of being both in the treatment group and in the post-period (i.e., the effect directly attributable to the program), is captured by the coefficients β_8, β_9 and β_{10}. The model interacts account participation in the treatment group during the post-period with weather through coefficients β_9 and β_{10}.

B.5 Estimate Savings

Navigant's model allowed for the calculation of savings to be represented as total savings, using Average Treatment Effects (ATE), and heating and cooling reduction, in units of savings per heating or cooling load. These representations are described in the following subsections.

B.5.1 Estimating Savings – Average Treatment Effect

Navigant estimated total savings using TMY3 data and ATE, representing energy savings in absolute values for typical weather years. To estimate the savings, Navigant calculated average treatment effects for variables that included participant usage during the post-period. To calculate annualized ATE, Navigant multiplied coefficient estimates for β_8, β_9, and β_{10} by the annualized mean value associated with those variables.[25] For example, coefficient β_9 had a mean HDD per year value of 4,964, which resulted in an annual ATE of -132, or a 132 kWh reduction in average annual usage relative to controls after correcting for weather. Table B-2 presents an overview of these calculations.

Table B-2. Calculating Annualized Average Treatment Effect

Coefficient	Treatment Effect Model Term	Estimate	Mean Value	ATE (kWh/year)
β_8	Savings per Day	0.31 (kWh/day)	365 (days/year)	114
β_9	Savings per HDD	-0.03 (kWh/HDD)	4,964 (HDD/year)	-132
β_{10}	Savings per CDD	-0.07 (kWh/CDD)	927 (CDD/year)	-63

Source: ComEd billing and tracking data and Navigant team analysis.

Summing the ATE values in Table B-2 results in a combined annual ATE of -81 kWh, or an 81 kWh reduction in average annual usage relative to controls after correcting for weather. This ATE value corresponds to the annual savings of 81 kWh as shown in the entry for "All PY8 Advanced Thermostat

[25] Weather coefficients used average monthly HDD and CDD values calculated with TMY3 data.

Participants with Adequate Data" in Table 5-2. Advanced Thermostat Research Estimated Savings by HER Status
.

B.5.2 Estimating Savings – Heating and Cooling Percentage Reductions

Navigant also represented savings as percentage reductions in energy used for heating and cooling. The model provides savings per day, per HDD, and per CDD. Savings per HDD represent savings associated with energy used for heating while savings per CDD represent savings associated with energy used for cooling. As agreed to during the November 2017 touchpoint meeting, baseline heating and cooling loads were used to break out "per day" savings into per-HDD and per-CDD savings. To calculate the reduction percentages, annual savings were divided by the respective heating and cooling loads to find percentage reductions in energy used for heating and cooling.

APPENDIX C. SENSITIVITY ANALYSES

This appendix contains results of the sensitivity analyses Navigant conducted. Navigant tested the sensitivity of the results estimated with the PY8 participant group to various comparison test groups, model specifications, and alternate data processing approaches.

C.1 Sensitivity to Different Matched Comparison Groups

Navigant conducted sensitivity analyses using different matched comparison groups to provide an indication of the sensitivity of the results to any one matched comparison group. Table C-1 compares the matching and observations requirements for the PY8 participant group with those of three comparison test groups[26].

Table C-1. Overview of Matching and Observations Requirements in Sensitivity Tests

Matched Comparison Group	Sample Size - Treatment	Sample Size - Control	Matching Method	Observation Requirements	Calipers Used to Filter Bad Matches
PY8 Participant Group	10,105	9,407	Sum of squared difference	Participants had 10 or more observations in all periods. Controls had 8 or more observations in all periods.	No
Comparison Test Group 1	8,193	8,193	Sum of absolute value of difference	10 months in all periods.	Yes
Comparison Test Group 2	10,451	9,974	Sum of absolute value of difference	10 or more observations during matching. 8 or more observations in pre- and post-periods	No
Comparison Test Group 3	10,105	9,661	Average squared difference	10 or more observations in all periods	No

Source: ComEd billing and tracking data and Navigant team analysis.

As shown in Table C-2, results from each of the comparison test groups are statistically similar to the PY8 participant group results. Furthermore, the PY8 participant group and each of the comparison test groups yielded savings estimates consistently lower than the values defined in the IL TRM. Navigant continued to make data improvements up until April 25, 2018, when the PY8 participant group was formed, and provides additional results from the comparison test groups as a robustness check on the findings. These robustness checks indicate that the findings are stable and are not dependent on a single comparison group.

[26] These groups correspond to the matched comparison groups named by date that were presented to evaluation research stakeholders. Comparison test groups 1 through 3 correspond to the 2018-02-06, 2018-04-03, and 2018-05-01 matched comparison groups

Table C-2. Sensitivity Analysis Results Using Different Matching and Observation Thresholds

Matched Comparison Group	Cooling		Heating		Total
	Percent Reduction (90% CI)	Electric Energy Savings by kWh/year/site	Percent Reduction (90% CI)	Electric Energy Savings by kWh/year/site	Electric Energy Savings by kWh/year/site (90% CI)
PY8 Participant Group	-0.3% (-2.0 to 1.4%)	-7	6.4% (4.4 to 8.4%)	88	81 (29 to 134)
Comparison Test Group 1	0.9% (-1.1 to 2.9%)	20	9.1% (6.7 to 11.4%)	116	136 (75 to 196)
Comparison Test Group 2	-1.7% (-3.2 to -0.3%)	-38	5.7% (3.8 to 7.6%)	79	41 (-12 to 94)
Comparison Test Group 3	-1.9% (-3.5 to -0.2%)	-41	5.7% (3.8 to 7.7%)	79	39 (-13 to 90)

Source: ComEd billing and tracking data and Navigant team analysis.

In Sections C.1.1 through C.1.4, Navigant provides additional information associated with the results of each comparison test group sensitivity analysis as well as the same information for the PY8 participant group for reference. For both the PY8 participant group and the comparison test groups, Navigant provides the regression model output, plots of average monthly usage for participants and controls in the matching and pre-treatment periods, and the numerical values used to create the plots.

C.1.1 PY8 Participant Group

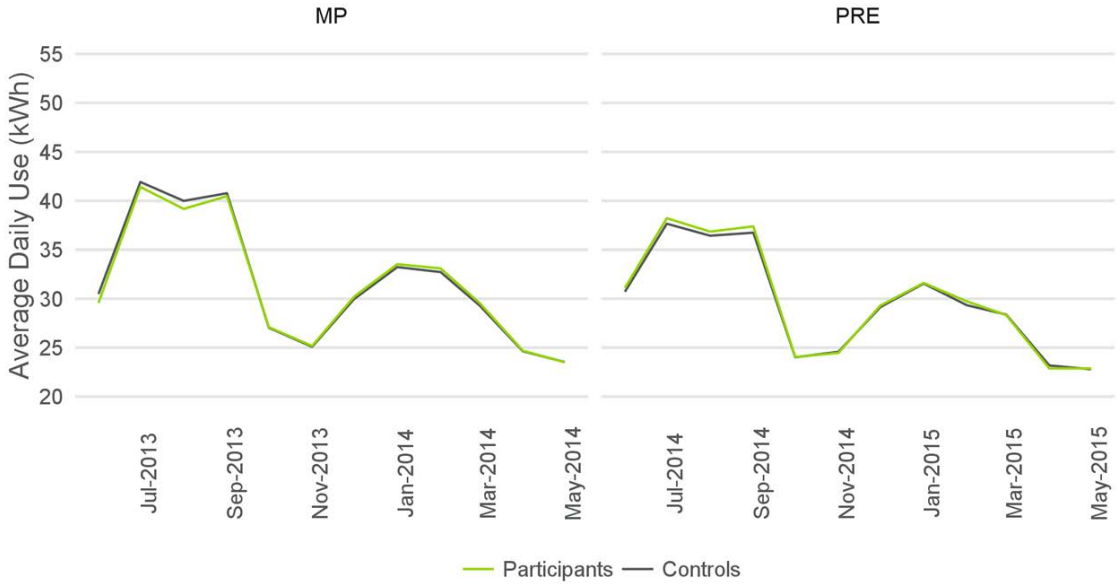
The PY8 participant group used data from participants that had 10 months of data in each period and controls that had eight or more months of data in each period. Participants and controls were matched using the sum of squared difference in energy usage.

Table C-3. Regression Output – PY8 Participant Group

Variable	Estimate	Standard Error	P-Value	Annual Electric Energy Use (kWh/year/site)
HDD	0.18	0.004	<0.01	911
CDD	1.64	0.017	<0.01	1,525
Post	-3.41	0.085	<0.01	-1,243
HDD*Post	0.09	0.003	<0.01	434
CDD*Post	0.62	0.014	<0.01	570
HDD*Treat	0.01	0.005	0.18	35
CDD*Treat	0.09	0.024	<0.01	80
Post*Treat	0.31	0.122	0.01	114
HDD*Post*Treat	-0.03	0.005	<0.01	-132
CDD*Post*Treat	-0.07	0.019	<0.01	-63

Source: ComEd billing and tracking data and Navigant team analysis.

Figure C-1. Matching Plot – PY8 Participant Group



Source: ComEd billing and tracking data and Navigant team analysis.

Table C-4. Matching Plot Data – PY8 Participant Group

Month	MP – 2013/2014 (kWh / day / site)			Pre-Period – 2014/205 (kWh / day / site)		
	Participants	Comparison	Difference	Participants	Comparison	Difference
6	29.58	30.48	-0.90	31.10	30.71	0.39
7	41.41	41.92	-0.51	38.22	37.66	0.56
8	39.17	39.99	-0.82	36.85	36.43	0.42
9	40.47	40.77	-0.30	37.39	36.74	0.65
10	27.09	27.04	0.06	24.04	24.02	0.03
11	25.17	25.09	0.08	24.47	24.58	-0.11
12	30.20	29.96	0.24	29.31	29.15	0.16
1	33.52	33.23	0.30	31.61	31.55	0.06
2	33.09	32.72	0.37	29.72	29.33	0.39
3	29.56	29.30	0.26	28.34	28.39	-0.05
4	24.68	24.63	0.06	22.88	23.18	-0.30
5	23.50	23.53	-0.03	22.88	22.79	0.09

Source: ComEd billing and tracking data and Navigant team analysis.

C.1.2 Comparison Test Group 1

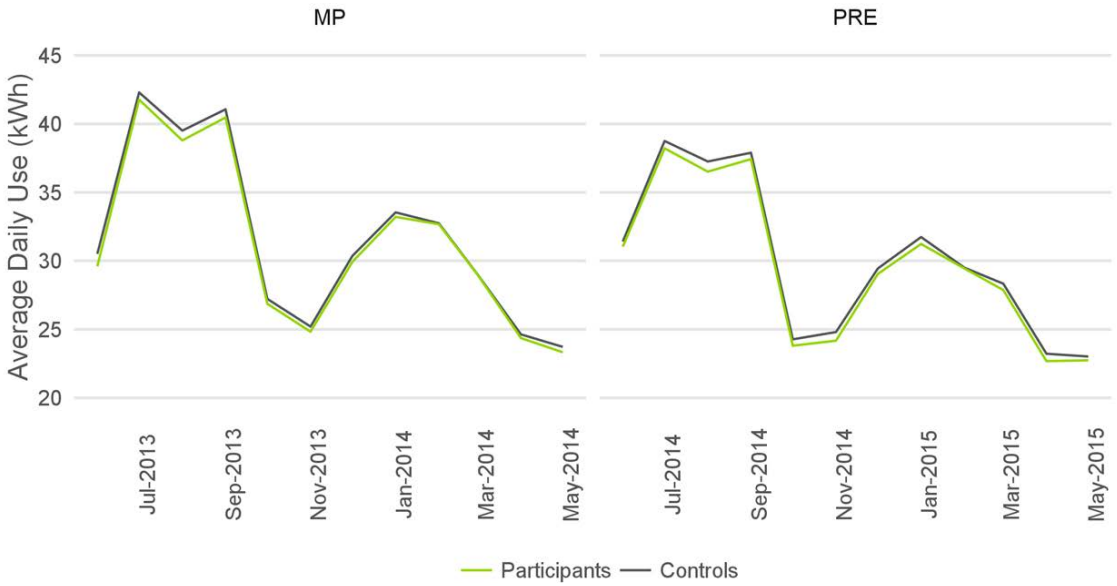
Comparison test group 1 used data from participants and controls that had 10 months of data in each period and were matched using the sum of absolute value of difference in usage.

Table C-5. Sensitivity Analysis Regression Output – Comparison Test Group 1

Variable	Estimate	Standard Error	P-Value	Annual Electric Energy Use (kWh/year/site)
HDD	0.18	0.004	<0.01	902
CDD	1.74	0.018	<0.01	1,614
Post*	-3.07	0.097	<0.01	-1,121
HDD*Post	0.07	0.004	<0.01	372
CDD*Post	0.60	0.015	<0.01	555
HDD*Treat	0.00	0.005	0.71	10
CDD*Treat	-0.01	0.025	0.73	-8
Post*Treat	0.23	0.134	0.09	83
HDD*Post*Treat	-0.03	0.005	<0.01	-147
CDD*Post*Treat	-0.08	0.020	<0.01	-71

Source: ComEd billing and tracking data and Navigant team analysis.

Figure C-2. Sensitivity Analysis Matching Plot – Comparison Test Group 1



Source: ComEd billing and tracking data and Navigant team analysis.

Table C-6. Sensitivity Analysis Matching Plot Data – Comparison Test Group 1

Month	MP – 2013/2014 (kWh / day / site)			Pre-Period – 2014/205 (kWh / day / site)		
	Participants	Comparison	Difference	Participants	Comparison	Difference
6	29.60	30.51	-0.91	31.04	31.41	-0.37
7	41.77	42.30	-0.53	38.21	38.75	-0.54
8	38.80	39.51	-0.71	36.52	37.26	-0.74
9	40.47	41.06	-0.60	37.43	37.89	-0.46
10	26.87	27.22	-0.35	23.81	24.28	-0.47
11	24.83	25.20	-0.38	24.17	24.81	-0.63
12	29.93	30.37	-0.43	29.05	29.45	-0.40
1	33.21	33.54	-0.33	31.23	31.74	-0.51
2	32.68	32.74	-0.07	29.45	29.51	-0.06
3	28.97	28.98	-0.02	27.87	28.34	-0.46
4	24.36	24.63	-0.27	22.68	23.22	-0.54
5	23.33	23.72	-0.39	22.74	23.02	-0.28

Source: ComEd billing and tracking data and Navigant team analysis.

C.1.3 Comparison Test Group 2

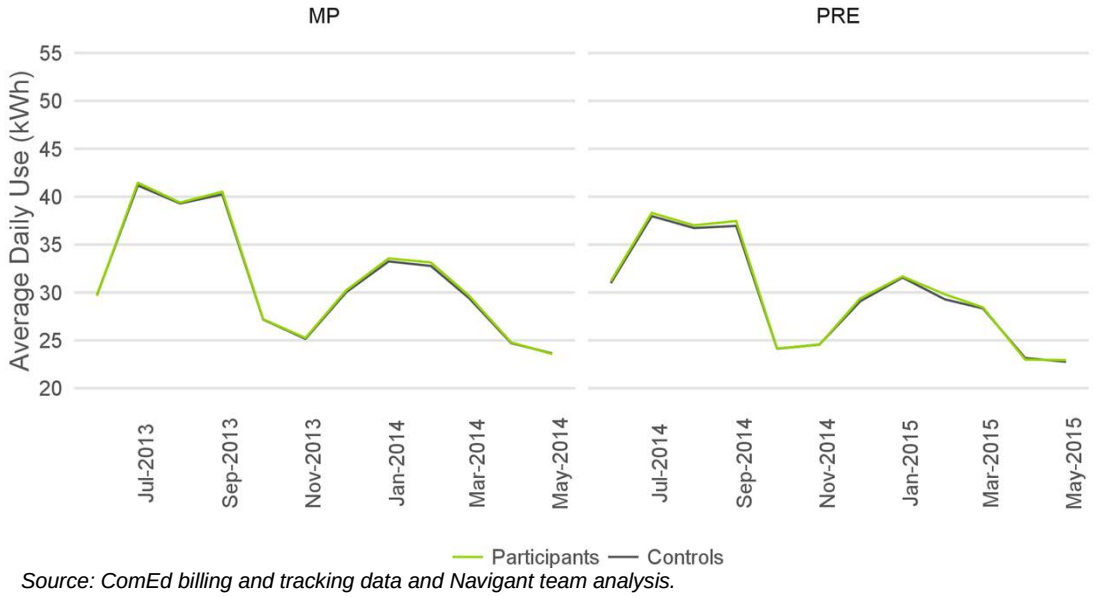
Comparison test group 2 used data from participants and controls that had 10 months of data in the matching period and eight or more months of data in the pre-treatment and post-treatment periods. Participants and controls were matched using the sum of absolute value of difference in energy usage.

Table C-7. Sensitivity Analysis Regression Output – Comparison Test Group 2

Variable	Estimate	Standard Error	P-Value	Annual Electric Energy Use (kWh/year/site)
HDD	0.18	0.004	<0.01	900
CDD	1.68	0.016	<0.01	1,559
Post*	-3.54	0.082	<0.01	-1,292
HDD*Post	0.09	0.003	<0.01	436
CDD*Post	0.60	0.013	<0.01	556
HDD*Treat	0.01	0.005	0.09	43
CDD*Treat	0.05	0.024	0.03	47
Post*Treat	0.40	0.120	<0.01	146
HDD*Post*Treat	-0.03	0.004	<0.01	-136
CDD*Post*Treat	-0.06	0.019	<0.01	-52

Source: ComEd billing and tracking data and Navigant team analysis.

Figure C-3. Sensitivity Analysis Matching Plot – Comparison Test Group 2



Source: ComEd billing and tracking data and Navigant team analysis.

Table C-8. Sensitivity Analysis Matching Plot Data – Comparison Test Group 2

Month	MP – 2013/2014 (kWh / day / site)			Pre-Period – 2014/205 (kWh / day / site)		
	Participants	Comparison	Difference	Participants	Comparison	Difference
6	29.66	29.71	-0.05	31.19	30.97	0.21
7	41.45	41.18	0.28	38.30	37.97	0.33
8	39.39	39.29	0.09	37.01	36.73	0.28
9	40.52	40.24	0.28	37.46	36.95	0.51
10	27.20	27.17	0.03	24.15	24.13	0.02
11	25.25	25.16	0.09	24.54	24.56	-0.02
12	30.24	30.05	0.20	29.38	29.09	0.30
1	33.55	33.24	0.31	31.67	31.56	0.11
2	33.13	32.77	0.36	29.81	29.28	0.52
3	29.66	29.41	0.25	28.44	28.33	0.11
4	24.78	24.71	0.07	22.99	23.15	-0.16
5	23.57	23.63	-0.06	22.93	22.77	0.17

Source: ComEd billing and tracking data and Navigant team analysis.

C.1.4 Comparison Test Group 3

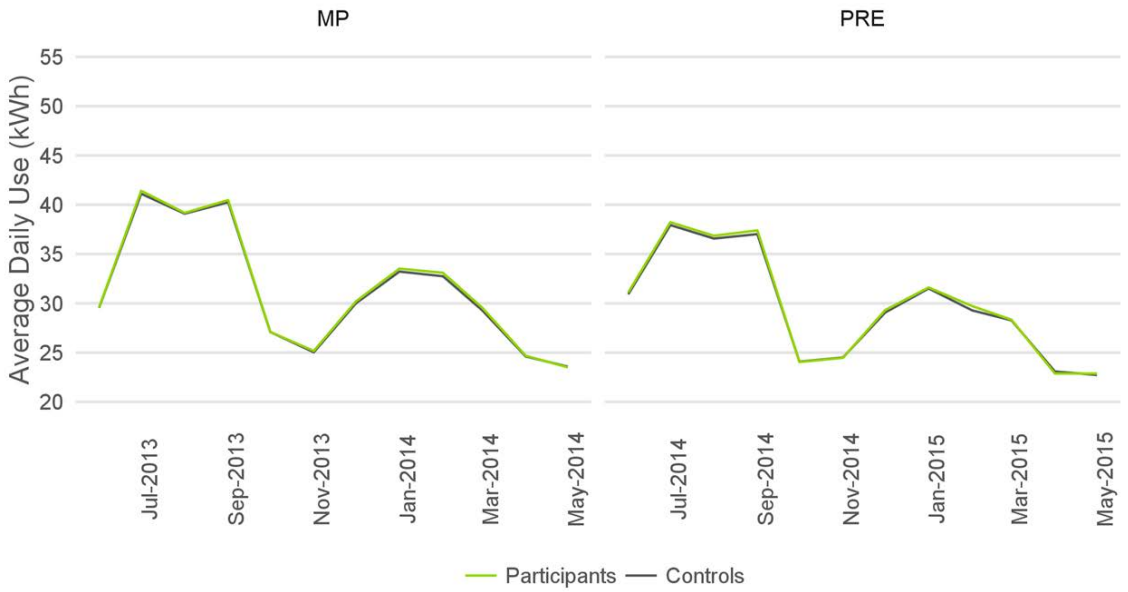
Comparison test group 3 used data from participants and controls that had 10 months of data in each period and were matched using the average squared difference in energy usage.

Table C-9. Sensitivity Analysis Regression Output – Comparison Test Group 3

Variable	Estimate	Standard Error	P-Value	Annual Electric Energy Use (kWh/year/site)
HDD	0.18	0.004	<0.01	898
CDD	1.67	0.017	<0.01	1,551
Post	-3.51	0.083	<0.01	-1,281
HDD*Post	0.09	0.003	<0.01	440
CDD*Post	0.60	0.013	<0.01	559
HDD*Treat	0.01	0.005	0.07	48
CDD*Treat	0.06	0.024	0.02	53
Post*Treat	0.42	0.121	<0.01	152
HDD*Post*Treat	-0.03	0.004	<0.01	-139
CDD*Post*Treat	-0.06	0.019	<0.01	-52

Source: ComEd billing and tracking data and Navigant team analysis.

Figure C-4. Sensitivity Analysis Matching Plot – Comparison Test Group 3



Source: ComEd billing and tracking data and Navigant team analysis.

Table C-10. Sensitivity Analysis Matching Plot Data - Comparison Test Group 3

Month	MP – 2013/2014 (kWh / day / site)			Pre-Period – 2014/205 (kWh / day / site)		
	Participants	Comparison	Difference	Participants	Comparison	Difference
6	29.58	29.55	0.03	31.10	30.91	0.19
7	41.41	41.11	0.30	38.22	37.93	0.29
8	39.17	39.09	0.09	36.85	36.57	0.28
9	40.47	40.25	0.22	37.39	37.02	0.38
10	27.09	27.09	0.00	24.04	24.08	-0.04
11	25.17	25.04	0.14	24.47	24.51	-0.04
12	30.20	30.01	0.19	29.31	29.07	0.24
1	33.52	33.22	0.30	31.61	31.51	0.10
2	33.09	32.74	0.35	29.72	29.28	0.44
3	29.56	29.29	0.27	28.34	28.27	0.06
4	24.68	24.61	0.07	22.88	23.09	-0.21
5	23.50	23.57	-0.07	22.88	22.73	0.15

Source: ComEd billing and tracking data and Navigant team analysis.

C.2 Sensitivity to Model Specification

To ensure model specification accuracy, Navigant conducted sensitivity analyses using different model specifications with the PY8 participant group. Table C-11 identifies the various model specifications used in the sensitivity analyses and Table C-12 presents the results of these analyses. None of the model

specifications yielded a higher savings estimate than the total savings for the TRM-recommended participant group (excludes HER effects) shown in Table 5-2. Advanced Thermostat Research Estimated Savings by HER Status
; the savings Navigant is recommending be used to guide updates to the IL TRM.

Table C-11. Robustness Model Specifications

Model	Robustness Check
PY8 participant group	Provided for comparison, no robustness check performed.
Simple Model	Did not include weather in model and used dummy variables and the comparison group to control for non-program effects on energy use
Weather Squared	Added squared HDD and CDD terms to the model
HER Binary	Used a binary dummy variable to identify HER recipients in the post-period
No Post*Treat	Did not include the post*trt interaction variable
No Weather*Post*Treat	Did not include weather (HDD or CDD) interacted with post times treatment

Source: Navigant team analysis.

Table C-12. Sensitivity Analyses Using PY8 Participant Group

Model	Average Daily Usage (kWh)	Average Customer Savings (kWh/yr)	Average Customer Savings Standard Error (kWh/yr)
PY8 participant group	30.35	81.45	31.81
Simple Model	30.35	66.17	32.23
Weather Squared	30.35	58.06	32.02
HER Binary	30.35	81.46	32.59
No Post*Treat	30.35	100.87	29.95
No Weather*Post*Treat	30.35	86.59	32.09

Source: ComEd billing and tracking data and Navigant team analysis.

C.3 Sensitivity to Alternate Data Processing

Navigant conducted a sensitivity analysis to determine the impacts of alternate data processing on estimated savings. In this analysis, weather data was merged into the PY8 participant group analysis dataset by bill date instead of by year and month. Table C-13 presents the estimated savings and Table C-14 presents the regression output for this sensitivity analysis. Estimated savings are not statistically different from those presented in Table 5-2. Advanced Thermostat Research Estimated Savings by HER Status
 at the 90% confidence level.

Table C-13. Sensitivity to Alternate Data Processing – Savings Estimates for the PY8 Participant Group with Weather Data Matched by Bill Date

Analysis	Cooling		Heating		Total
	Reduction (%) (90% CI)	Electric Energy Savings by kWh/year/site	Reduction (%) (90% CI)	Electric Energy Savings by kWh/year/site	Electric Energy Savings by kWh/year/site (90% CI)
Evaluation Research Results Group, Matches Weather Data by Bill Date	-0.02% (-1.3 to 1.3%)	-1	4.4% (2.9 to 5.9%)	88	88 (36 to 140)

Source: ComEd billing and tracking data and Navigant team analysis.

Table C-14. Sensitivity to Alternate Data Processing – Regression Output for the PY8 Participant Group with Weather Data Matched by Bill Date

Variable	Estimate	Standard Error	P-Value	Annual Electric Energy Use (kWh/year/site)
HDD	0.27	0.00	<0.01	1,359
CDD	2.08	0.02	<0.01	1,926
Post	-3.47	0.09	<0.01	-1,265
HDD*Post	0.11	0.00	<0.01	567
CDD*Post	0.50	0.01	<0.01	464
HDD*Treat	0.01	0.01	0.02	73
CDD*Treat	0.13	0.03	<0.01	116
Post*Treat	0.48	0.12	<0.01	175
HDD*Post*Treat	-0.03	0.00	<0.01	-166
CDD*Post*Treat	-0.10	0.02	<0.01	-97

Source: ComEd billing and tracking data and Navigant team analysis.

APPENDIX D. ADDITIONAL MATERIALS SUPPORTING ILLINOIS TRM UPDATES

This appendix contains content related to requests from ICC staff meant to provide further detail on the results of the evaluation research and inform updates to the IL TRM.

D.1 Signs of Self-Selection Bias Prior to Program Participation

Navigant searched for diverging trends between comparison and treatment groups before advanced thermostats were installed to investigate whether there were indicators of self-selection bias. Navigant implemented this analysis by using the regression model to compare energy usage impacts between the matching and pre-treatment periods as opposed to between the pre-treatment and post-treatment periods. Navigant conducted this test for both the PY8 participant group and comparison test group 3.

Table D-1 and Table D-2 show the results of this analysis for the PY8 participant group and comparison test group 3, respectively. As shown in the tables, there are two indications that there were not diverging trends during the two years before advanced thermostats were installed. The first indication is that the post treatment variable coefficient values are smaller than what are shown by the regression results in Table C-3. This is most clear in the HDD*Post*Treat term coefficient value; while the coefficient value in Table C-3 is relatively small, an impact is observable. In Table D-1 and Table D-2, the coefficient values for this term are zero indicating that there is no program effect during these time periods. Second, the p-values for the post treatment variables are high, indicating that these terms are not statistically significant. This is in contrast with the statistically significant treatment terms in Table C-3. Subsequently, we see that participants and control customers used energy similarly up until the program intervention.

Table D-1. Time Period Comparison – PY8 Participant Group Regression Output

Variable	Estimate	Standard Error*	P-Value*	Annual Electric Energy Use (kWh/year/site)
HDD	0.28	0.004	<0.01	1,414
CDD	2.18	0.017	<0.01	2,024
Post	2.39	0.072	<0.01	872
HDD*Post	-0.10	0.003	<0.01	-505
CDD*Post	-0.54	0.013	<0.01	-501
HDD*Treat	0.01	0.005	0.15	39
CDD*Treat	0.05	0.025	0.03	49
Post*Treat	-0.12	0.106	0.27	-42
HDD*Post*Treat	0.00	0.004	0.85	-3
CDD*Post*Treat	0.03	0.019	0.10	29

* Error is inaccurate when including the matching period in the regression dataset. Navigant expects the error provided here is biased low, but does not have any indication on the magnitude.
Source: ComEd billing and tracking data and Navigant team analysis.

Table D-2. Time Period Comparison – Comparison Test Group 3 Regression Output

Variable	Estimate	Standard Error*	P-Value*	Annual Electric Energy Use (kWh/year/site)
HDD	0.28	0.004	<0.01	1,398
CDD	2.20	0.016	<0.01	2,039
Post	2.28	0.070	<0.01	832
HDD*Post	-0.10	0.002	<0.01	-500
CDD*Post	-0.53	0.012	<0.01	-489
HDD*Treat	0.01	0.005	0.04	55
CDD*Treat	0.04	0.024	0.12	34
Post*Treat	-0.01	0.104	0.96	-2
HDD*Post*Treat	0.00	0.004	0.66	-8
CDD*Post*Treat	0.02	0.018	0.29	18

* Error is inaccurate when including the matching period in the regression dataset. Navigant expects the error provided here is biased low, but does not have any indication on the magnitude.
Source: ComEd billing and tracking data and Navigant team analysis.

D.2 Results for Non-HER Participants in Comparison Test Group 3 Using Various Models

This subsection contains regression outputs from various models for comparison test group 3, filtered to advanced thermostat participants who did not participate in the HER program. These results were found using the model with the post*treatment term (i.e., the agreed upon model) and a model without the post*treatment term included.

D.2.1 Summary of Comparison Test Group 3, Non-HER Estimated Savings Using Various Models

Table D-3 presents a comparison of the results from a model with and a model without the post*treatment term for non-HER participants in comparison test group 3. Removing the post*treatment term leads to increased cooling savings which are statistically different than zero at the 90% confidence level. To comply with best practices, Navigant supports referencing 2% cooling reduction as the value most applicable from this study to inform IL TRM updates, because it reflects the approach agreed to prior to seeing the analysis results. Conducting analysis under this framework increases transparency and the repeatability of results.

Table D-3. Comparison Test Group 3, Non-HER Participant Estimated Savings Between Models

Model	Cooling		Heating		Total
	Reduction (%) (90% CI)	Electric Energy Savings (kWh/year/site)	Reduction (%) (90% CI)	Electric Energy Savings (kWh/year/site)	Electric Energy Savings (kWh/year/site) (90% CI)
With Post*Treatment	1.7% (-1.8 to 5.3%)	34	6.0% (1.7 to 10.4%)	77	111 (7 to 214)
Without Post*Treatment	3.2% (0.9 to 5.5%)	61	5.3% (0.5 to 10.2%)	66	58 (-20 to 135)

Source: ComEd billing and tracking data and Navigant team analysis.

D.2.2 Comparison Test Group 3, Non-HER Participant Characteristics and Data Processing

In preparation for this analysis, Navigant processed non-HER participant data by removing customers and data points from the analysis in the steps identified in Table D-4.

Table D-4. Comparison Test Group 3, Non-HER Participant Site Attrition

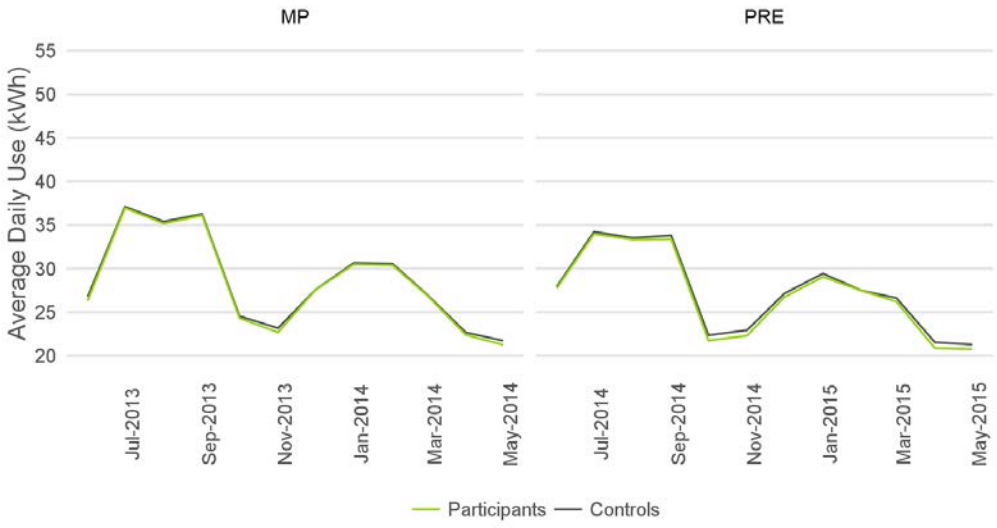
Data Processing Step	Participants	Controls
Raw Data	23,884	1,946,260
Filter to non-HER participants	10,202	275,409
Bad Reads (e.g., All Estimated Bills)	10,202	275,408
Averaged Bills in the Same Month	10,202	275,408
Removed Outliers (i.e., observations one order of magnitude above or below median usage)	10,202	275,383
Had 10+ Observations in Pre- and Post-Periods	3,932	186,474
Received No Other Rebates or Rebated Measures from Res HVAC or HEA Programs	3,399	179,226
Had Matches in same HER Wave and Zip	3,399	164,173
Had 10+ Observations in the Matching Period	2,641	155,610

Source: ComEd billing and tracking data and Navigant team analysis.

D.2.3 Comparison Test Group 3, Non-HER Participant Matching

This subsection contains information on the participant matching for comparison test group 3, non-HER participants. Figure D-1. Comparison Test Group 3, Non-HER Participant Matching Plot presents a plot of average monthly usage for participants and controls in the matching and pre-treatment periods and Table D-5 presents the numerical values used to create the plot.

Figure D-1. Comparison Test Group 3, Non-HER Participant Matching Plot



Source: ComEd billing and tracking data and Navigant team analysis.

Table D-5. Comparison Test Group 3, Non-HER Participant Matching Data

Month	MP – 2013/2014 (kWh / day / site)			Pre-Period – 2014/205 (kWh / day / site)		
	Participants	Comparison	Difference	Participants	Comparison	Difference
6	26.31	26.77	-0.46	27.74	27.93	-0.19
7	36.94	37.10	-0.17	33.96	34.23	-0.27
8	35.17	35.39	-0.22	33.33	33.53	-0.20
9	36.08	36.26	-0.18	33.38	33.80	-0.42
10	24.32	24.55	-0.23	21.74	22.36	-0.62
11	22.69	23.17	-0.48	22.30	22.93	-0.63
12	27.52	27.55	-0.03	26.74	27.15	-0.41
1	30.48	30.65	-0.17	29.05	29.43	-0.38
2	30.38	30.56	-0.18	27.46	27.48	-0.01
3	26.90	26.99	-0.09	26.22	26.63	-0.42
4	22.39	22.67	-0.28	20.87	21.54	-0.67
5	21.25	21.72	-0.48	20.76	21.29	-0.53

Source: ComEd billing and tracking data and Navigant team analysis.

D.2.4 Comparison Test Group 3, Non-HER Regression Output

Table D-6 provides regression output from the model with the post*treatment term and Table D-7 presents regression output from the model without the post*treatment term.

Table D-6. Comparison Test Group 3, Non-HER Participant Regression Output – With Post*Treatment Term

Variable	Estimate	Standard Error	P-Value	Annual Electric Energy Use (kWh/year/site)
HDD	0.17	0.009	<0.01	856
CDD	1.40	0.033	<0.01	1,295
Post*	-2.74	0.167	<0.01	-1,000
HDD*Post	0.07	0.006	<0.01	366
CDD*Post	0.63	0.026	<0.01	585
HDD*Treat	0.01	0.012	0.31	62
CDD*Treat	0.07	0.047	0.11	69
Post*Treat	0.27	0.242	0.26	99
HDD*Post*Treat	-0.02	0.009	0.01	-116
CDD*Post*Treat	-0.10	0.036	0.01	-93

Source: ComEd billing and tracking data and Navigant team analysis.

Table D-7. Comparison Test Group 3, Non-HER Participant Regression Output – Without Post*Treatment Term

Variable	Estimate	Standard Error	P-Value	Annual Electric Energy Use (kWh/year/site)
HDD	0.17	0.009	<0.01	868
CDD	1.41	0.033	<0.01	1,304
Post*	-2.60	0.121	<0.01	-951
HDD*Post	0.07	0.006	<0.01	341
CDD*Post	0.61	0.024	<0.01	569
HDD*Treat	0.01	0.012	0.52	40
CDD*Treat	0.05	0.048	0.25	51
HDD*Post*Treat	-0.01	0.008	0.08	-66
CDD*Post*Treat	-0.07	0.030	0.03	-61

Source: ComEd billing and tracking data and Navigant team analysis.

D.3 Analysis of Heating and Cooling Loads Between Comparison Groups

Navigant also conducted an analysis around heating and cooling loads between participants and comparison group customers before advanced thermostats were installed. Table D-8 and Table D-9 contain weather variable regression output for the PY8 participant group and comparison test group 3, respectively. These results indicate a difference of approximately 4% of heating and cooling use between comparison and treatment groups before households received advanced thermostats. The savings model controls for this difference in weather sensitivity as shown in the results for the HDD*treat and CDD*treat variables.

Table D-8. PY8 Participant Group Matches Weather Variable Regression Output

Variable	Estimate	Standard Error	P-Value	Annual Electric Energy Use (kWh/year/site)
HDD	0.24	0.00	<0.01	1,188
CDD	1.93	0.02	<0.01	1,785
HDD*Treat	0.01	0.01	0.08	44
CDD*Treat	0.08	0.02	<0.01	78

Source: ComEd billing and tracking data and Navigant team analysis.

Table D-9. Comparison Test Group 3 Matches Weather Variable Regression Output

Variable	Estimate	Standard Error	P-Value	Annual Electric Energy Use (kWh/year/site)
HDD	0.24	0.00	<0.01	1,181
CDD	1.96	0.02	<0.01	1,820
HDD*Treat	0.01	0.01	0.04	51
CDD*Treat	0.05	0.02	0.04	43

Source: ComEd billing and tracking data and Navigant team analysis.

APPENDIX E. STAKEHOLDER COMMENTS

Navigant received comments on this report in draft form from Ecobee and Nest. The draft was posted August 27, 2018 and comments were received on or before September 11, 2018. Navigant appreciates all stakeholder comments, and includes them in this report's appendix to ensure they are documented for the public. Stakeholder comments are helpful in understanding the research results, and subsequently informing the TRM, a process led by VEIC. Navigant provides in the sections below a summary of the comments, a summary of Navigant's responses, and the comments with Navigant's detailed responses imbedded.

E.1 Summary of Comments Received

Navigant summarized and highlighted comments we believe to be most relevant in Table E-1 but both Ecobee and Nest provided more detail in their submitted comments, which can be found in section6.E.3

Table E-1. Summary of Stakeholder Comments

Received Comments	Ecobee	Nest Labs	Navigant's Response Subsection
Disagree with results of current study and recommend not updating TRM until new methodologies and results can be verified	✓	✓	Suggested IL TRM Updates
Concerned about self-selection bias and/or participant groups not being representative	✓	✓	Modeling Issues
Concerned about changes in baseload between pre- and post-installation periods	✓	✓	
Concerned about customer matching	✓		
Suggests additional research questions and methods that could improve understanding and accuracy of estimated savings values, including leveraging AMI data	✓	✓	Future Research

Source: Navigant's summary and highlights of comments received from Ecobee and Nest.

E.2 Summary of Navigant's Response to Comments

Navigant summarizes its response to comments in the sections below.

E.2.1 Suggested IL TRM Updates

Stakeholders expressed concern with Navigant's recommendation to reference 2% as the results from this study most applicable for informing the IL TRM, until further research is conducted. Navigant notes that ultimately the IL TRM Administrator and the IL TRM TAC are responsible for IL TRM updates. The purpose of this report is to document the research conducted by Navigant (using methods agreed to by the Advanced Thermostat Subcommittee). Navigant supports additional research that would better inform data-driven updates to the IL TRM, with consideration of available evaluation resources.

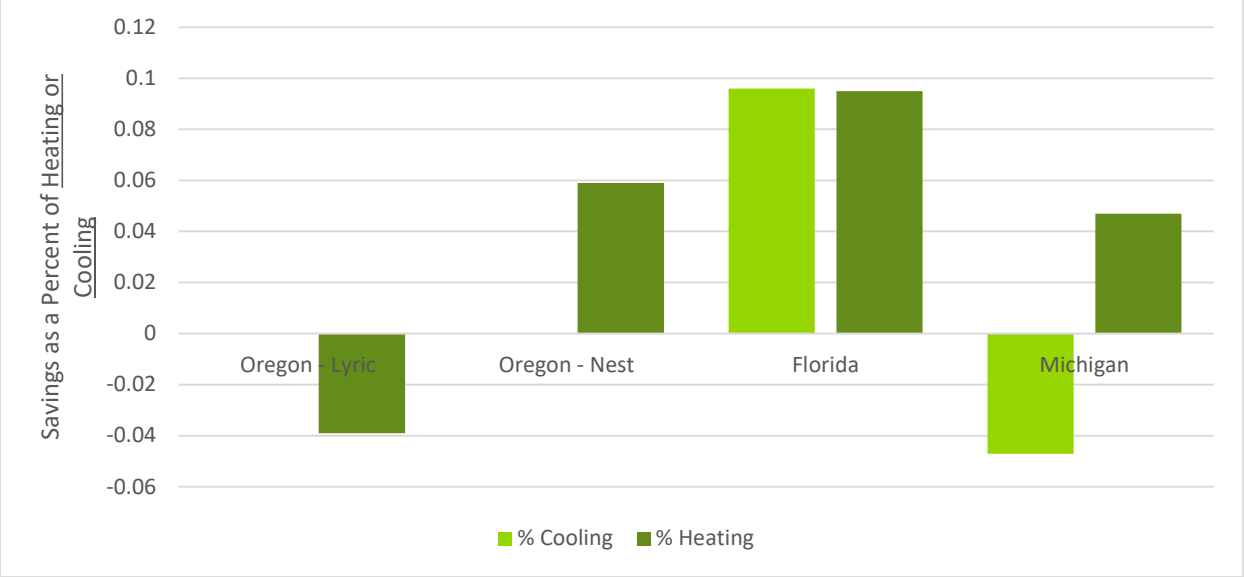
Additionally, it may not be possible to evaluate this measure such that there is absolutely no uncertainty in the results. Typically, the IL TRM TAC uses the best available information as a part of data driven IL TRM updates. Navigant proposes for the reader to be aware that:

- Navigant's approach and the results are in line with other third-party evaluations across the US. See additional details below under "Secondary Literature."
- Other possible methods for evaluation lean on assumptions that are no more reliable than assumptions made in this analysis. See additional details below under "Other Evaluation Approaches."
- This measure could be considered for custom evaluation rather than for the IL TRM. See additional details below under "IL TRM or Custom."

Secondary Literature

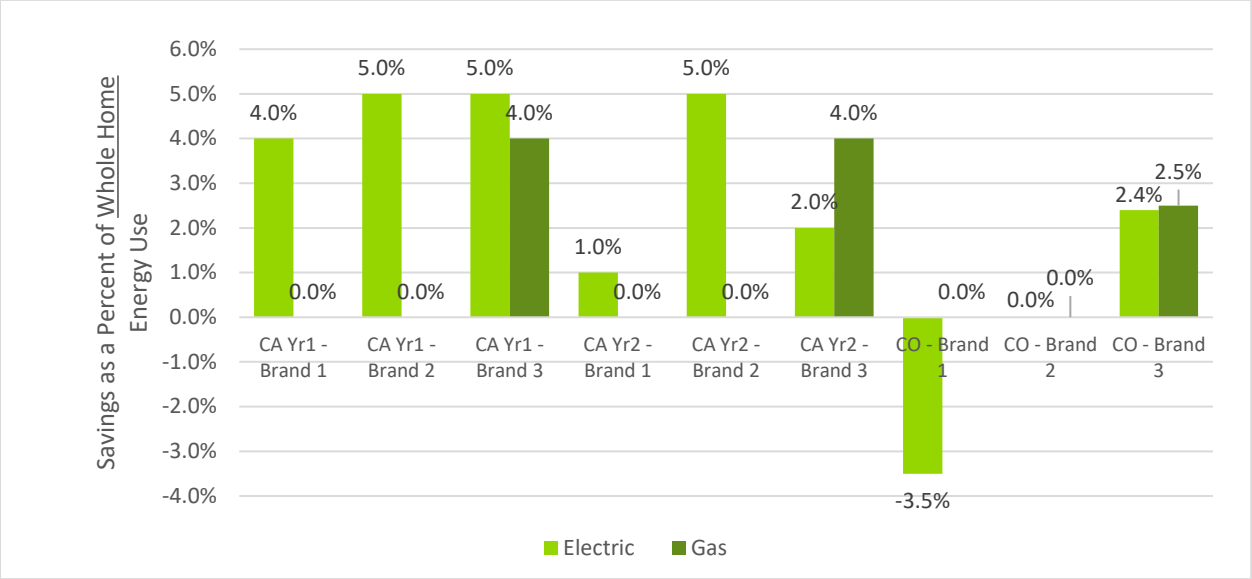
Navigant documents in this subsection third party evaluations of advanced thermostat savings outside of IL published in 2016 or later. These evaluations used methods similar to those employed in IL. We separate results by those calculated as a percent of heating or cooling load and those calculated as a percent of whole home energy use. Navigant made an effort to include as many studies as possible, but acknowledges that our list is not exhaustive. In summary, Navigant interprets the results as exemplifying the variability of advanced thermostat savings. In general, when savings vary between regions and within regions, Navigant warns against using results from other regions and applying them to IL. Of note, in a hot climate, one evaluation found zero heating savings. While not clear evidence, this finding in combination with the findings in IL (clear heating savings, but low cooling savings) might be indicative that advanced thermostats save less energy in mild climates where participants may utilize on/off HVAC behavior in the baseline.

Figure E-1. Third Party Evaluation Findings where Savings are Presented per Heating or Cooling Energy Use



Source: Navigant's summary of secondary literature shared with VEIC July 2018 as a part of the IL TRM redline process.

Figure E-2. Third Party Evaluation Findings where Savings are Presented per Whole Home Energy Use



Source: Navigant's summary of secondary literature shared with VEIC July 2018 as a part of the IL TRM redline process.

Other Evaluation Approaches

All evaluations require some assumptions for interpreting the results for a TRM. The assumptions used in Navigant's evaluation are arguably more reliable than other possible approaches.

For lighting evaluations, evaluators typically meter hours-of-use after the measure is installed, ultimately making the assumption that the hours-of-use are consistent before and after the measure is installed. For advanced thermostats, we know this assumption is not appropriate, because the measure is intended to reduce runtime. Metering could be useful for this measure if evaluators could meter participants before installing advanced thermostats. This effort would require substantial evaluation resources and would lean on two key assumptions: (1) the metered sample of customers perfectly represents the participant population, and (2) all trends affecting energy use (e.g., weather and the economy) are stable enough between the pre and post-installation period for the evaluation team to properly estimate savings. Metering is a reasonable approach to evaluation but uses assumptions that are no more reliable than the assumptions used in this analysis.

Without metering, the ENERGY STAR metric leverages data from advanced thermostats to approximate savings. However, because data is not readily available regarding the baseline, the ENERGY STAR metric assumes that customers used flat, stable internal temperatures prior to installing advanced thermostats. However, there are multiple sources of evidence that this assumption does not hold for Illinois.

First, as discussed at the May 8, 2018 touch-point meeting, about 50% of PY8 participants in this analysis reported that they used automatic setbacks with programmable thermostats. In ComEd's 2013 baseline study (conducted by Opinion Dynamics), 67% of customers reported that they programmed their programmable thermostats. Although survey responses are not as reliable as other sources of data, this information indicates that customers in IL use setbacks (at least to some extent) even in the absence of advanced thermostats.

Additionally, Opinion Dynamics and Cadmus conducted a 2013 submetering study of manual and programmable thermostats during the cooling season in IL and found that customers used a variety of efficient control strategies. They found that:[27]

- Only 18% of participants did not adjust manual thermostats regularly.
- Manually controlled thermostats are turned off for longer periods of time than programmable thermostats.
- Homes using automatic setbacks may use less efficient set points when occupied. Homes with manually controlled thermostats have higher average indoor temperatures. Homes with regularly scheduled temperature settings have an average indoor temperature lower than those who do not.
- Constant temperature set point homes tolerate higher average temperatures and turn off the HVAC system for longer periods of time than those following a program.

In summary, there is no method for evaluating this measure where the uncertainty is zero. In working with the subcommittee to develop this method, Navigant feels the method mitigated assumptions to the greatest extent possible given data availability and evaluation resources. Navigant looks forward to continuing to work with the Advanced Thermostat Subcommittee in developing future research that best balances evaluation resources, rigor and insightfulness.

IL TRM or Custom

This measure could be considered for custom evaluation as opposed to being included in the IL TRM. If Ecobee and Nest feel that the methods available, that are used in other evaluations in the US, are not suitable for evaluating this measure in IL, Navigant suggests that Ecobee and Nest consider proposing (1) that this measure be removed from the IL TRM and (2) that the program run the measure as a randomized experiment, as is done with HER. Navigant feels that moving this measure to custom and running a randomized program like HER would put unnecessary burden on programs. However, Navigant also does not feel comfortable accepting this measure in the IL TRM without evaluation.

E.2.2 Modeling Issues

Some key issues raised by stakeholders are discussed in this section. All stakeholders commented on a variety of possible modeling issues that may have influenced results. In general, Navigant appreciates committee member feedback and notes that this study involved a thorough analysis of the sensitivity of results to several factors, including matching methodology and model specification. Navigant found that these different modeling scenarios estimated similar impacts. Similarly, the methods employed for this study are commonly used to evaluate energy efficiency programs and measures, including for advanced thermostats. For example, advanced thermostat evaluation studies for Xcel Energy and in MI used similar methods to Navigant's most recent evaluation of advanced thermostats in IL.[28] [29] As a result, although individual model parameters may have a degree of uncertainty, the consistency of the savings estimates across different and commonly used models corroborates the findings from this study: the electric savings value for these IL-specific participants does not support the cooling reduction factor in the IL TRM version 6.

Self-Selection Bias and Non-Representative Participant Groups

Nest Labs commented that the results of this study show that the participant groups exhibited evidence of self-selection bias that would influence results. Navigant notes that self-selection bias is a problem for all

[27] Cadmus and Opinion Dynamics, 2013. "Thermostat Assessment for Cooling Savings."

[28] Xcel Energy Evaluation: https://www.xcelenergy.com/staticfiles/xe-responsive/Company/Rates%20&%20Regulations/Regulatory%20Filings/CO-Smart-Thermostat-Pilot-Evaluation.PDF

[29] MI Evaluation Research: https://www.michigan.gov/documents/mpsc/Tier_3_Tstat_Calibration_Study_EWR_Presentation_623038_7.pdf

studies that do not use a randomized control trial. Consequently, the methodology of this study was designed to handle the issue of self-selection bias as rigorously as possible, including customer matching, comparison of multiple matching groups, and use of regression models to test for signs of self-selection bias before program participation, where no signs of self-selection bias were found. Navigant presents more details on the sensitivity analysis in 6.Appendix C and 6.Appendix D. While these additional analysis outputs don't prove the absence of self-selection, they do serve as additional rigor beyond standard practice, and show no signs of self-selection bias. These results indicate that if self-selection bias is affecting results, the effect would have to be equal across the various comparison groups and the effect would have to start occurring simultaneously with the program, having not been occurring prior.

Changes in Energy Use for Non-Participants
Ecobee and Nest expressed concern regarding the large changes in energy use between the pre- and post-installation periods among the comparison groups. Navigant was not able to conduct any analysis to understand non-program related trends in energy consumption, but will include this research for consideration in future evaluations. With regards to savings, Navigant tested seven different model specifications and used four different comparison groups, which all yielded similar findings. If there are concerns associated with certain parameters in certain models, Navigant notes that the issue would have to be consistently present across all models and comparison groups for it to compromise the findings from this study.

Participant "Baseload" Changes over Time
Ecobee and Nest suggest that Navigant exclude the "per day" savings (i.e., post*treatment) from the advanced thermostat savings estimate. Navigant was not able to conduct any analysis to better understand this finding, but will include this research for consideration in future evaluations. However, with regards to this study, Navigant compromised with a thermostat manufacturer to include this parameter (the "main treatment effect") in the estimate for savings in order to avoid underestimating savings.[30] Changing the interpretation of this variable after seeing results is not appropriate. Furthermore, although this finding is unexpected, this variable is used as standard practice to estimate savings in other evaluations, in industry guidance documents, and in academic literature.

E.2.3 Future Research

Navigant appreciates the suggestions that were submitted for future research that could improve savings estimates and is willing to consider them for future research. Navigant proposes to estimate annual cooling savings leveraging the following data sources if they can be made available: ComEd advanced metering infrastructure (AMI) data, Ameren Illinois AMI data (where available), a survey of participants, and advanced thermostat telemetry data. Navigant will also consider additional research for future evaluations around this measure's effect on baseload and non-program related changes in energy consumption. Navigant will coordinate this research with the Advanced Thermostat Subcommittee. When considering the scope of this research, Navigant suggests the group balance the needs of this study compared to other evaluation needs in the region given limited evaluation resources. Additionally, if the subcommittee is unable to reach consensus on future evaluation results, Navigant proposes that the IL TRM TAC consider removing this measure from the IL TRM and using custom evaluation (as is done with HER) – there should not be measures in the IL TRM where the savings are so nuanced that they cannot be evaluated.

[30] A manufacturer commented on December 13, 2016 "the proposed regression model (Equation 1) estimates savings only through an interaction term with cooling degree days (CDD70) and does not include any main treatment effect" and the subcommittee agreed on January 20, 2017 during the touch-point meeting to include this additional variable in the model.

E.3 Comments Received

In this section, Navigant provides the stakeholder comments in black text with Navigant's responses imbedded in green text.

Ecobee Comments:

We disagree with the two percent cooling reduction recommendation in the evaluation research report for the same reasons that we have expressed since Navigant released the evaluation's regression outputs in May 2018. These concerns have not been lessened by the additional analyses Navigant conducted to address uncertainties surrounding the results. Our concerns with the report are as follows:

- Implausible regression results: The data presented in Table C-3 of the report clearly shows unusual results between the pre- and post-periods. The post-period exhibits a large reduction in energy while there is a simultaneous large increase in usage on a per degree day basis.
  Navigant was not able to conduct any analysis to better understand non-program related trends in energy consumption and can only offer speculation on this topic. With regards to savings, Navigant conducted the analysis using a model agreed upon by the subcommittee as a reasonable approach to estimate savings. Navigant tested seven different model specifications and used four different comparison groups, which all yielded similar findings. If there are concerns associated with certain parameters in certain models, Navigant notes that the issue would have to be consistently present across all models and comparison groups for it to compromise the findings from this study. Furthermore, and given the consistency in savings estimates across the sensitivity testing, the results from this study are likely the most applicable research findings for informing the IL TRM. With this information, the IL TRM administrator can choose to incorporate these findings with other findings in updates to the IL TRM as determined through the IL TRM TAC.
  As we highlighted previously, while the treatment group shows savings on a per degree day basis, the savings are effectively eliminated by the other terms. These inconsistencies have yet to be explained.
  The available data indicates that this measure affects energy use uncorrelated with heating or cooling. Navigant was not able to conduct any analysis to better understand this finding, but will include this research for consideration in future evaluations. As explained in earlier responses to comments and within this report, Navigant compromised with a thermostat manufacturer to include this parameter (the "main treatment effect") in the estimate for savings in order to avoid underestimating savings.[31] Changing the interpretation of this variable after seeing results is not appropriate. Furthermore, although this finding is unexpected, this variable is used as standard practice to estimate savings in other evaluations, in industry guidance documents, and in academic literature.
  We still question how Navigant undertook matching for the regression analysis – and the final report only serves to raise more questions about this element of the evaluation.
  For matching, Navigant matched participants to non-participants with the most similar pre-installation energy use within the same HER wave and zip code. We've held several meetings to discuss the matching and responded to previous comments from Ecobee on the matching process. In the future, if there are lingering questions, please reach out for explanation prior to reporting. Additionally, if there are any specific questions, please let us know and we'd be happy to discuss them with you.

[31] A manufacturer commented on December 13, 2016 "the proposed regression model (Equation 1) estimates savings only through an interaction term with cooling degree days (CDD70) and does not include any main treatment effect" and the subcommittee agreed on January 20, 2017 during the touch-point meeting to include this additional variable in the model.

For example, it is not clear how Navigant addressed the missing months in either the treatment or control group.
Matching was conducted on only months where participants and prospective matches each had data.

- Wide confidence intervals: Contrary to the report's conclusions, we believe the wide confidence interval ranges support our continued rejection of the results. For example, on Table 5-2, the confidence intervals are large enough that significant savings could still have been realized.
  Wide confidence intervals do not invalidate results. They only identify the level of uncertainty in estimating smart thermostat savings, which is critical information for informing decision making. The more variable the savings are for a measure, the wider the confidence interval will be. It is worth noting that Navigant conducts monthly billing analysis for other EE programs with similar participant sizes and finds smaller confidence intervals. Furthermore, the results of this study (even at the upper error bounds) are statistically lower than the IL TRM v6.0 specifies, indicating that some adjustment to the IL TRM is warranted. Just as it is true that these results indicate that the true cooling savings may be higher than 2%, it is equally likely that they are lower, possibly being negative.

Overall, we agree that the method for evaluating advanced thermostats must change going forward – particularly now that a sufficient number of ComEd customers have smart meters installed. As we and others recommended, the new method should incorporate AMI data to determine the setback behavior of customers who do not have smart thermostats.

We support future research using AMI. While it is not clear that AMI data will inherently lead to a different estimate of savings, AMI data could lead to improved error bounds and give us a clearer picture of heating and cooling loads.

Thank you,
Nkechi Ogbue
Manager, Regulatory Affairs

Nest Comments:

Thank you for the opportunity to comment; we raise questions/challenges with the completed evaluation study and propose near- and longer-term opportunities to improve understanding of Illinois smart thermostat savings. We look forward to working in partnership with the other Illinois stakeholders to address these challenges and to move forward together.
Navigant looks forward to coordinating on future research as well. Thanks!

<u>Questions/Challenges with the Study</u>

The current smart thermostat billing data analysis study was designed to evaluate the electric savings attributable to installing smart thermostats for customers who participated in ComEd's thermostat rebate program between June 2015 and May 2016. For any study, there are questions about how best to interpret the statistical findings and then how much weight to give those findings in potentially updating savings values in the TRM.

The study involved analyzing the monthly electric billing data for a subset of participants before and after the program year and included a comparison group to try to capture any trends in electric usage unrelated to the program. The comparison group was selected from other ComEd customers in the same zip codes as participants and with comparable prior electric usage patterns. The monthly electric usage data was then analyzed using a linear regression

model designed to account for weather variations over time and estimate the net savings attributed to the thermostats.

For any study of an on-going program, there are always questions about each of three key aspects of the evaluation:

1. Self-selection Bias: How confident can we be that the comparison group accurately reflects the changes in electric use that the participants would have experienced without the program?
2. Regression Modeling: Does the regression model reliably capture usage patterns and accurately identify the incremental changes for participants?
3. Participant group: Were the participants included in the analysis representative of the full participant population? And, for TRM purposes, are they representative of future participants to whom the TRM value will be applied?

Different answers to these questions can lead to very different interpretations of the same study. We present a summary of the interpretation put forth by Navigant and then provide some alternative explanations for these findings and order the topics from most to least important.

Self-Selection Bias

The study assumes that the matched comparison group perfectly represents how electric usage would have changed over time for the participants if they had not participated.
This assumption is absolutely critical for the study to yield accurate results. If the comparison and participant groups' energy usage would have diverged even without the program, then the results of the study are invalid.
All evaluations require some assumptions for interpreting the results for a TRM. The assumption identified is used very commonly in evaluation, as it is often considered a more reliable assumption than the assumptions required for other approaches. Furthermore, based on previous feedback, Navigant conducted additional analysis and used multiple comparison groups to test this assumption to the greatest extent possible. Navigant's additional testing did not lead to any evidence refuting the comparison group's representativeness. While this concern is warranted, the only path forward to eliminate this concern entirely would be to remove this measure from the IL TRM and run the program as a randomized experiment, as is done with HER.

The customers in the treatment group -- customers who bought a smart thermostat -- selected themselves into the treatment group raising the potential for self-selection bias. Customers who buy smart thermostats may differ from other customers in ways we don't know. This may be especially true for early adopters. The matching process is intended to reduce this bias as much as possible but one can never be sure how much bias remains.

Smart thermostat technology is unlikely to appeal to all demographic groups equally. Research has found that smart thermostat buyers are generally younger and wealthier than the general population.[32]
This discontinuity further demonstrates the importance of using evaluation for this measure. Without evaluation-supported EE benefits, one could argue that this measure serves as a mechanism to transfer wealth from general ComEd customers to the "younger and wealthier" ComEd customers.

[32] see " New York Smart Thermostat Market Characterization", by Northeast Energy Efficiency Partnerships and SEE Change Institute. NYSERDA Oct. 2016
https://www.nyserda.ny.gov/-/media/Files/Publications/PPSER/Program-Evaluation/2016ContractorReports/Smart-Thermostat-Market-Charaterization-Report.pdf

These differences can result in customers on different trajectories in terms of their home energy usage. For example, if young couples with a positive economic outlook are more likely to buy a smart thermostat then participants may be more likely to experience changes such as having a baby and purchasing additional energy-consuming technology and also may be less likely to become empty nesters during the study period. There may also be differences in terms of working from home, embarking on major home renovations, and other lifestyle shifts. These differences between smart thermostat buyers and the comparison group could easily amount to 2% or more of annual electric use -- which is as large as the expected savings from the smart thermostats.

Matching is the best available approach to account for these trends, albeit, imperfectly. Based on previous feedback, Navigant conducted additional analysis and used multiple comparison groups to test this assumption to the greatest extent possible. Navigant's additional testing did not lead to any evidence refuting the comparison group's representativeness. While this concern is warranted, the only path forward to eliminate this concern entirely would be to remove this measure from the IL TRM and run the program as a randomized experiment, as is done with HER.

Self-selection bias can easily go undetected, although some assessments can be made using the available data and through customer surveys. In this instance, the report shows strong evidence of self-selection bias in two ways: 1) usage patterns of the two groups diverge before the smart thermostat were purchased and;

Based on previous feedback, Navigant conducted additional testing and found that the two groups were not diverging in a measurable way before advanced thermostats were installed. Please see Appendix 6.Appendix C and 6.Appendix D. While there is variation in energy use on average, Navigant ran the model to estimate savings during the periods prior to measure installation and found no evidence that the groups were diverging. Additionally, some of the comparison groups showed almost no variation in energy use during the matching period and pre-installation period and these comparison groups showed the same consistent results – the PY8 participant electric savings are statistically lower than the IL TRM v6.0 specifies.

2) baseload usage among participants apparently increased significantly from a measure that does not affect baseload.

The available data indicates that this measure does affect energy use uncorrelated with heating or cooling. Navigant was not able to conduct any analysis to better understand this finding, but will include this research for consideration in future evaluations. Navigant compromised with a thermostat manufacturer to include this parameter (the "main treatment effect") in the estimate for savings in order to avoid underestimating savings.[33] Changing the interpretation of this variable from savings to self-selection bias after seeing results is not appropriate. Furthermore, although this finding is unexpected, this variable is used as standard practice to estimate savings in other evaluations, in industry guidance documents, and in academic literature.

First, the report has data that shows a clear and significant increase in cooling season electric use for the participants vs. the comparison group before the program. The comparison group was selected to match the participants' electric usage for the period June 2013 through May 2014 (the matching year). The most critical assumption of the evaluation is that this matching of 2013/14 data creates a comparison group that will accurately reflect how participant electric usage would have changed all the way through May 2017 (the end of the post period).

[33] A manufacturer commented on December 13, 2016 "the proposed regression model (Equation 1) estimates savings only through an interaction term with cooling degree days (CDD70) and does not include any main treatment effect" and the subcommittee agreed on January 20, 2017 during the touch-point meeting to include this additional variable in the model.

Based on previous feedback, Navigant conducted additional testing and found that the two groups were not diverging in a measurable way before advanced thermostats were installed. Please see Appendix 6.Appendix C and 6.Appendix D. While there is variation in energy use on average, Navigant ran the model to estimate savings during the periods prior to measure installation and found no evidence that the groups were diverging, i.e., no evidence of savings from advanced thermostats before they were installed. Additionally, some of the comparison groups showed almost no variation in energy use during the matching period and pre-installation period and these comparison groups showed the same consistent results – the PY8 participant electric savings are statistically lower than the IL TRM v6.0 specifies.

The report also includes monthly usage data for the two groups for the period June 2014 through May 2015 (the pre-program year). Any divergence would mean this critical assumption is violated and the analysis is not valid. The data show a large shift in the cooling season electric usage of the participants vs. the comparison group. The shift amounts to a 139 kWh increase in cooling season usage for the participants -- a 9% increase. The comparison group's failure to track the participant group means that the key assumption of the analysis doesn't hold.

Based on previous feedback, Navigant conducted additional testing and found that the two groups were not diverging in a measurable way before advanced thermostats were installed. Please see 6.Appendix C and 6.Appendix D. Our testing used regression models (exactly those used to estimate savings), which is more rigorous and more applicable to understanding this issue than using simple averages.

Perhaps an easier way to understand the implications of this finding is that, if one agrees with the study's conclusion that there was no self-selection bias, then one should conclude that buying a smart thermostat a year in the future increased cooling usage by 9% before buying it. This conclusion does not make sense. This shift is clear evidence of substantial self-selection bias.

Based on previous feedback, Navigant conducted additional testing and found that the two groups were not diverging in a measurable way before advanced thermostats were installed. Our regression model showed 0 effect of the thermostat on pre-installation energy use, i.e., zero savings from a future smart thermostat purchase, rather than 9%. Please see 6.Appendix D.

In response to this problem being pointed out, Navigant shared results from alternative versions of the matching which didn't show such large obvious bias and that resulted in similar savings estimates. But just because other iterations of the matching did not all show such clear evidence of bias does not mean that there is not any bias -- just that it is not always so obvious. The main conclusion still stands -- that smart thermostat program participants are different from randomly selected customers who live nearby and have similar usage patterns in 2013/14 and so self-selection bias is a major problem with this evaluation.

The comparison group was not selected randomly, but based on neighbors with similar energy usage patterns. Self-selection bias is a challenge, but all evaluations require some assumptions for interpreting the results for a TRM. The assumption around matching is used very commonly in evaluation, as it is often considered a more reliable assumption than the assumptions required for other approaches. Furthermore, based on previous feedback, Navigant conducted additional analysis and used multiple comparison groups to test this assumption to the greatest extent possible. Navigant's additional testing did not lead to any evidence refuting the comparison group's representativeness. While this concern is warranted, the only path forward to eliminate this concern entirely would be to remove this measure from the IL TRM and run the program as a randomized experiment, as is done with HER.

The second piece of evidence indicating self-selection bias is that the analysis found a significant increase in baseload electric usage for participants -- but smart thermostats are not

expected to affect baseload usage. The report claims that the reported 114 kWh increase in baseload usage was caused by the smart thermostats because it led participants to buy other smart home technology.

This statement is not relevant to this report. Navigant does not suggest in this report that the 114 kWh is from smart home technologies. Navigant was not able to conduct any analysis to better understand this issue and can only offer speculation. Interactive effects from smart home technologies was one speculation offered in previous meetings and was mentioned as possibly contributing to the 114 kWh/year/participant, rather than fully explaining it.

That explanation is not only unlikely (the causation could go the other way or one underlying factor could cause both) but also can't realistically account for the average increase of 114 kWh. Every participant would need to buy 4-5 smart home products to increase baseload usage by that much (and the comparison group would need to buy 0) -- a very unlikely scenario. A more reasonable explanation for this increase in baseload usage is self-selection bias -- that participants were following a different trend in energy use due to other differences in their life circumstances.

The available evidence suggests that this measure leads to negative savings uncorrelated with heating or cooling. Navigant was not able to conduct any analysis to better understand this finding, but will include this research for consideration in future evaluations. Navigant compromised with a thermostat manufacturer to include this parameter (the “main treatment effect”) in the estimate for savings in order to avoid underestimating savings.[34] Changing the interpretation of this variable from savings to self-selection bias after seeing results is not appropriate. Furthermore, although this finding is unexpected, this variable is used as standard practice to estimate savings in other evaluations, in industry guidance documents, and in academic literature.

It's worth noting that only the baseload portion of self-selection bias would be directly noticeable in the results because the baseload usage shouldn't change. But the evidence of bias from increasing baseload usage implies that there may be bias in the cooling and heating loads as well. For example, if the self-selection effect is that participants are increasing their daytime occupancy (e.g., having more children) then it's not just baseload usage that would have increased but also heating and cooling usage. But the increases in heating and cooling loads would be hidden within the observed heating and cooling savings estimates as simply lower savings estimates.

The available evidence suggests that this measure leads to negative savings uncorrelated with heating or cooling. Navigant was not able to conduct any analysis to better understand this issue and can only offer speculation. For one possible explanation, previous metering studies suggest that some ComEd customers leave their thermostat off more often when not using setbacks.[35] As a result, advanced thermostat participants may use more cooling in mild weather once they have an advanced thermostat, because they choose to leave on their cooling more often. These negative savings could show up in the model as “baseload” if they occur during the shoulder season when heating and cooling degree days are minimal.

The very real possibility that self-selection bias could result in a 1% or 2% bias in total usage is why billing data analysis is not usually used to evaluate savings from measures with small expected impacts. For example, savings from lighting retrofits are typically estimated based on field studies using light loggers, not from billing data analysis.

[34] A manufacturer commented on December 13, 2016 “the proposed regression model (Equation 1) estimates savings only through an interaction term with cooling degree days (CDD70) and does not include any main treatment effect” and the subcommittee agreed on January 20, 2017 during the touch-point meeting to include this additional variable in the model.

[35] Cadmus and Opinion Dynamics, 2013. “Thermostat Assessment for Cooling Savings.”

Navigant uses monthly billing data to evaluate several programs that have 1% to 2% savings including Home Energy Reports and Bidgely. Additionally, metering would only be useful for this measure if evaluators could meter participants before installing advanced thermostats. This effort would require substantial evaluation resources and would lean on two key assumptions: (1) the metered sample of customers perfectly represents the participant population, and (2) all trends affecting energy use (e.g., weather and the economy) are stable enough between the pre and post-installation period for the evaluation team to properly estimate savings. Metering is a reasonable approach to evaluation but uses assumptions that are no more reliable than the assumptions used in this analysis. If Nest feels the savings of this measure are too nuanced to be evaluated using standard approaches, Nest should consider proposing that the program be run as a randomized experiment, as is done with HER.

Regression Modeling

The study assumes that the regression model accurately captured the heating and cooling dynamics of these homes and provided accurate estimates of heating, cooling, and baseload savings.

This statement is false. Navigant conducted matching and sensitivity analysis on the model specification to avoid leaning solely on this assumption. Across the various comparison groups and model specifications, the electric savings are statistically lower than the IL TRM v6.0 specifies. The approach used is aimed at making as few and as reliable of assumptions as possible. In causal inference analysis, odd coefficients that are unrelated with savings are not expected to compromise the findings.

Based on the regression outputs, the following statements describe what the model showed about the comparison group's usage between the pre and post periods:

- average cooling usage increased by 37% per degree day -- meaning after accounting for weather,
- average heating usage increased by 48% per degree day,
- average baseload usage dropped by 1,243 kWh/year, and
- total weather adjusted usage declined by 239 kWh -- reflecting offsetting increases in HVAC and decreases in baseload.

The explanation for the huge baseload usage decline was that homes are becoming more efficient over time due to naturally occurring efficiency improvements in appliances and lighting, but this explanation does not account for how heating increased by 48% or how cooling increased by 37%.

Navigant was not able to conduct any analysis to understand non-program related trends in energy consumption and can only offer speculation on this topic. With regards to savings, Navigant conducted the analysis using a model agreed upon by the subcommittee as a reasonable approach to estimate savings. Navigant tested seven different model specifications and used four different comparison groups, which all yielded similar findings. If there are concerns associated with certain parameters in certain models, Navigant notes that the issue would have to be consistently present across all models and comparison groups for it to compromise the findings from this study. Furthermore, and given the consistency in savings estimates across the sensitivity testing, the results from this study are likely the most applicable research findings for informing the IL TRM. With this information, the IL TRM administrator can choose to incorporate these findings with other findings in updates to the IL TRM as determined through the IL TRM TAC.

The regression model also reported that the thermostats produced statistically significant cooling and heating savings but these savings were largely offset by an increase in baseload. Specifically, the thermostats saved 195 kWh in HVAC usage -- 63 kWh in cooling and 132 kWh of heating -- but the baseload usage increased by 114 kWh. This baseload increase was then

allocated to the cooling and heating loads which is what created the zero estimated cooling savings. This allocation is based on the assumption that the thermostat caused the baseload usage to increase. If it really did cause the baseload usage to change, then perhaps the TRM should include separate baseload "savings" for smart thermostats and not allocate it to cooling and heating.
Navigant proposed this option to the IL TRM administrator May 24, 2018, i.e., to update the IL TRM to allow for a baseload savings from this measure separate from heating or cooling.

Regression Modeling: Alternative View

The huge estimated shifts in heating, cooling, and baseload usage for the comparison group simply don't make sense. The claim that naturally occurring efficiency improvements could explain the drop in annual baseload use of 1,243 kWh doesn't add up. Every home would have to make changes as large as completely unplugging two refrigerators. The explanation of generally increasing efficiency would also make the huge heating and cooling usage increases even less sensible.
Navigant was not able to conduct any analysis to understand non-program related trends in energy consumption and can only offer speculation on this topic. Navigant will include this research for consideration in future evaluations. With regards to savings, Navigant conducted the analysis using a model agreed upon by the subcommittee as a reasonable approach to estimate savings. Navigant tested seven different model specifications and used four different comparison groups, which all yielded similar findings. If there are concerns associated with certain parameters in certain models, Navigant notes that the issue would have to be consistently present across all models and comparison groups for it to compromise the findings from this study. Furthermore, and given the consistency in savings estimates across the sensitivity testing, the results from this study are likely the most applicable research findings for informing the IL TRM. With this information, the IL TRM administrator can choose to incorporate these findings with other findings in updates to the IL TRM as determined through the IL TRM TAC.

The combination of a large drop in baseload coupled with huge increases in estimated heating and cooling intensity suggests that the regression model somehow under-estimated baseload in the post year which caused inflated estimates of heating and cooling use (less of the summer usage would be classified as baseload forcing the rest into cooling). The 239 kWh estimated decline in total usage may be reasonable and is consistent with residential usage trends. But the regression model did a poor job of estimating the break out of load shifts between heating, cooling, and baseload.

Another anomaly in the regression model results is that heating use was estimated at 1,345 kWh in the post treatment year. That value is about twice as large as the expected electric use of gas furnaces (which is what nearly every home in this analysis has).
Some participants have electric heat. This estimate aligns with the IL TRM estimate of heating load given the split between electric and gas heat.

It's clear that something is off with the regression model. The problem could be some sort of data error with the usage or weather data. Or perhaps it's just a reflection of the difficulty in accurately modeling heating and cooling usage from monthly utility bills in a climate like Illinois where non-HVAC loads account for the vast majority of electric use. Seasonal variations in lighting, refrigeration, and other end uses across the year can lead to heating and cooling estimates that don't make a lot of sense.

If the regression model estimates of the loads and load shifts for the comparison group don't

make sense, then it's hard to rely upon the same model to estimate how these load changes differed between the participant and comparison groups. For example, the model says that comparison group cooling increased by 37% while the participant cooling increased by just 32%. It's hard to have much faith in the accuracy of these estimates.
The approach used is aimed at making as few and as reliable of assumptions as possible. In causal inference analysis, odd coefficients that are unrelated with savings are not expected to compromise the findings. Navigant conducted the analysis using a model agreed upon by the subcommittee as a reasonable approach to estimate savings. Navigant tested seven different model specifications and used four different comparison groups, which all yielded similar findings. If there are concerns associated with certain parameters in certain models, Navigant notes that the issue would have to be consistently present across all models and comparison groups for it to compromise the findings from this study. Furthermore, and given the consistency in savings estimates across the sensitivity testing, the results from this study are likely the most applicable research findings for informing the IL TRM. With this information, the IL TRM administrator can choose to incorporate these findings with other findings in updates to the IL TRM as determined through the IL TRM TAC.

Even if one ignores these modeling problems and accepts the savings estimates as is, the results can still be interpreted quite differently. The first (HER + non-HER participants) regression estimated 195 kWh/yr in heating and cooling savings but that a 114 kWh increase in baseload wiped out most of these savings. But if the baseload increase was not caused by the thermostat purchase but was instead a reflection of self-selection bias, then the savings from the thermostats would be 195 kWh/yr. If self-selection bias caused the increase in baseload usage (e.g., through increased daytime occupancy), then the bias may also have increased the cooling and heating loads. But that portion of the bias would be hidden within the current heating and cooling estimates. So the actual total HVAC savings could be larger than the 195 kWh -- perhaps even larger than the current TRM values.
The available data indicates that this measure increases energy use uncorrelated with heating or cooling. Navigant was not able to conduct any analysis to better understand this finding, but will include this research for consideration in future evaluations. Navigant compromised with a thermostat manufacturer to include this parameter (the “main treatment effect”) in the estimate for savings in order to avoid underestimating savings.[36] Changing the interpretation of this variable from savings to self-selection bias after seeing results is not appropriate. Furthermore, although this finding is unexpected, this variable is used as standard practice to estimate savings in other evaluations, in industry guidance documents, and in academic literature.

Participant Group

The study assumes that the participant analysis group is perfectly representative of the participant population. Furthermore, in proposing to use the results "as-is" for the TRM, they are assuming that the savings from the 2015/16 program are the same as the savings for future participants.
Evaluations based on previous participants are used as standard practice to inform TRM’s. While this practice is imperfect, it enables the IL TRM to provide saving estimates that are based on the best available research. Alternatively, if this measure were removed from the IL TRM, programs would evaluate this measure annually, which would eliminate this concern. Navigant feels that moving this measure to custom would put unnecessary burden on programs, but also does not feel comfortable accepting this measure in the IL TRM without evaluation.

[36] A manufacturer commented on December 13, 2016 “the proposed regression model (Equation 1) estimates savings only through an interaction term with cooling degree days (CDD70) and does not include any main treatment effect” and the subcommittee agreed on January 20, 2017 during the touch-point meeting to include this additional variable in the model.

Participant Group: Alternative View

The participant analysis group excluded the majority of participants from the 2015/16 program year. The main reasons for exclusion were that the customer didn't have sufficient monthly utility data for the June 2013 - May 2017 period -- only customers with a long history at the same address were included. The slides included a few comparisons between the analysis and larger participant groups that revealed several differences. The most obvious difference being that the analysis group was more likely to report having an existing programmable thermostat that was actually running a program.
The analysis data set and the PY8 participant population self-reported that 51% and 44% respectively had programmed programmable thermostats in the baseline. Given the uncertainty in self-reporting, this difference is not large. Furthermore, Navigant found similar cooling savings for customers with manual and programmed programmable thermostats as the baseline, and as such, did not propose to adjust the cooling results based on baseline equipment.

But an even larger concern for using these results in the TRM is that this study analyzed a different program design -- where customers had to pay for the ~$250 thermostat and then later received a rebate. The current program includes less expensive thermostats and an instant rebate that reduces out of pocket costs to $69 for Nest's newest ENERGY STAR smart thermostat. In other words, the types of people in the analysis group may be very different from expected participants going forward and therefore savings for the current program may differ from those for the earlier program, limiting the value of results from the earlier design.
Evaluations based on previous participants are used as standard practice to inform TRM's. While this practice is imperfect, it enables the IL TRM to provide saving estimates that are based on the best available research. Alternatively, if this measure were removed from the IL TRM, programs would evaluate this measure annually, which would eliminate this concern. Navigant feels that moving this measure to custom would put unnecessary burden on programs, but also does not feel comfortable accepting this measure in the IL TRM without evaluation. If stakeholders feel the savings are so nuanced that commonly used evaluation approaches are not suitable, and as such, evaluation results for this measure should not be used to inform the IL TRM, Navigant suggests they consider proposing to remove this measure from the IL TRM and use a custom, randomized program design (as is done with HER).

Conclusions

Perhaps the biggest lesson from the current study is that it's hard to accurately measure smart thermostat savings when those savings are expected to be just a small fraction of the total electric bill. Self-selection bias could easily explain any difference between the current study and the TRM values. The increase in baseload usage and the matching problems provide evidence to support the self-selection explanation. The explanation for this, that smart thermostats somehow caused customers to increase their baseload usage, seems far less likely.
There is no evidence supporting the self-selection explanation, but there is no way to know for sure whether it is affecting the results. The groups were not diverging prior to advanced thermostats being installed (please see 6.Appendix D) and the change in baseload was originally proposed to capture "the main treatment effect" to avoid underestimating savings. There are unexpected findings and Navigant proposes to include research around those topics in future evaluations. Navigant recommends this measure either be included in the IL TRM and use the best information available given the agreed upon assumptions, or move this measure to custom evaluation (like HER) where the program design by necessity is randomized. If the stakeholders determine that the savings for this measure are too nuanced for common evaluation techniques, Navigant suggests they consider proposing to move this measure to custom. While Navigant feels that treating this measure as custom would put unnecessary burden on IL

programs, it is not appropriate to leave this measure in the IL TRM without evaluation, and all evaluation approaches require some assumptions.

In any case, the prudent course would seem to be to assess how confident anyone can be in the current study results, perform any additional analysis work needed to investigate these issues further, and then decide on what the study really shows and how reliable and precise those conclusions might be. The next step would be considering how best to incorporate any conclusions within the larger framework of potential TRM revisions based on the weight of evidence.
To update the IL TRM v7.0, the IL TRM administrator proposed to take a weighted average of relevant savings options. Navigant supported their conclusions and approach.

We believe that an analysis of available AMI (smart meter) data might be able to shed some light on some of these concerns. Specifically, the AMI data could be used to get a better assessment of changes in baseload usage and also help assess whether occupancy patterns were shifting for participants in different ways than the comparison group. Thermostat device data from manufacturers may also be useful in assessing some aspects of the modeling. An exploration of these datasets may shed light on some of these key unresolved issues.
Navigant agrees that AMI data (as well as thermostat device data from manufacturers) could provide additional insight into energy usage of participants and the comparison group.

To obtain Illinois specific data on setback behavior Nest recommends against relying solely on self-reporting surveys because people tend to over report behavior they perceive to be deemed "virtuous". Other approaches need to be used such as analysis of AMI data and potentially direct field observations of thermostat settings for a sample of participants in an existing direct install program. Nest remains committed to working with Illinois energy efficiency stakeholders to identify and implement emerging best practices in smart thermostat evaluation.
Navigant appreciates this feedback and will plan to coordinate with the Advanced Thermostat Subcommittee for future research. Navigant agrees that surveys are not perfectly reliable and supports using AMI data and thermostat device data from manufacturers. However, just like surveys, neither AMI data nor thermostat device data from manufacturers can provide perfectly clear data on baseline thermostat operation. Furthermore, participant surveys are used as standard practice in IL evaluation in estimating net-to-gross. Alternatively, sub-metering would only be useful for this measure if evaluators could meter participants before installing advanced thermostats. This effort would require substantial evaluation resources and would lean on two key assumptions: (1) the metered sample of customers perfectly represents the participant population, and (2) all trends affecting energy use (e.g., weather and the economy) are stable enough between the pre and post-installation period for the evaluation team to properly estimate savings. Metering is a reasonable approach to evaluation but uses assumptions that are no more reliable than the assumptions used in this analysis. For example, if a sub-metering effort were to come to an unexpected conclusion, stakeholders could argue that the unexpected finding is due to an unrepresentative sample rather than being a meaningful finding. In that case, evaluators would use substantial resources, while gaining no actionable information. In summary, there is uncertainty in all possible evaluation approaches and Navigant looks forward to coordinating with the Advanced Thermostat Subcommittee to find an approach to future research that best balances evaluation resources, rigor and insightfulness.

Sincerely,

Will Baker
Customer Success - Energy Partnerships | Google