

Bills, Bills, Bills: Can Utility Meter Data Tell the Heat Pump Story?

Patrick Hewlett, DNV

ABSTRACT

Heat pumps have become a critical component of several states' greenhouse gas emissions reduction plans due to their efficient offset of carbon-intensive fossil fuels. But assessing heat pump performance and impacts is challenging due to the broad array of displaced heating fuels, the possibility of partial displacement, and wide variation in usage patterns among customers. The authors of this paper have conducted an evaluation of three programs in the Northeast that rebated air-source and ground-source heat pump installations among primarily residential customers. The phased study has provided a unique opportunity to refine site-specific savings with increasing rigor—from initial at-the-meter analysis to equipment-specific measurement and verification (M&V) with yearlong performance data continually transmitted over cloud-communicating devices.

The phased approach invites comparison between the two methods. At-the-meter analysis is significantly cheaper than M&V, but can it explain why real-world savings may deviate from program claims? Conversely, equipment-level M&V is expensive, but does its higher rigor provide useful insights on how to improve heat pump savings forecasts? The authors explore these topics while comparing the results of each phase.

As heat pump adoption is forecasted to scale dramatically nationwide, utility program administrators are challenged with developing reliable, fuel-specific impact estimates that capture the wide range of heat pump usage patterns. This paper explores the best practices for scalable heat pump savings estimation and identifies the most appropriate methods to assess their progress versus goals.

Introduction

As utilities strive to meet aggressive decarbonization goals, space-conditioning heat pumps have become increasingly critical components of the energy transition. While heat pumps offer more control and modularity than traditional whole-home fossil fuel-fired systems, their uptake has been slower than desired due to barriers in capital costs, customer hesitation to part with legacy HVAC systems, and a reputation for suboptimal performance in extreme winter temperatures. Customers that have adopted ductless mini-split heat pumps (DMSHPs) in particular have often done so as a supplement to an existing heating system. This partial electrification has made predicting heat pump usage patterns and impacts difficult to predict. As a result, program administrators (PAs) have struggled to set realistic but aggressive goals, fine-tune incentives, target ideal customers, and develop a contractor base that advocates for heat pumps while satisfying customer needs.

To address many of those challenges, the authors of this paper have concluded a two-phase impact evaluation of three heat pump programs in the Northeast. The programs supported both residential and non-residential customers with the installation of ducted air-source heat pumps (ASHPs), DMSHPs, or ground-source heat pumps (GSHPs) to displace or replace preexisting systems that typically consumed fossil fuels or electric resistance for heating.

Program eligibility criteria did not require the preexisting heating or cooling equipment to be removed. 99% of rebated ASHPs were DMSHPs, all of which were classified as “cold climate.”

Evaluation Phase 1 involved analysis of site-specific, premise-level energy consumption at a sample of 235 heat pump installations. The analysis was supported by a web-based survey that provided basic information on facility characteristics, occupancy and usage patterns, and preexisting conditions. To ensure sufficient post-installation data for premise-level analysis using utility billing data, evaluators analyzed projects installed from program inception through 2018. Phase 1 provided a comparatively swift and economical assessment of premise-level heat pump impacts among regulated and unregulated fuels.

Evaluation Phase 2 involved equipment-level measurement and verification (M&V) at 137 sampled projects. Field engineers deployed cloud-communicating meters that transmitted operational data over a metering period of 12-18 months. During the site visit, field engineers conducted a more in-depth interview to confirm and refine information from the initial web survey: for example, what the customer would have done absent participation in the program. Phase 2 served as a more comprehensive and rigorous assessment of equipment-level impacts but required a more significant investment of time and budget.

This paper compares and contrasts the methods and results between Phase 1 and Phase 2. Throughout much of this paper, we use total MMBtu savings across all fuels (including electricity, at site¹) as a simplified proxy for total energy savings and subsequent carbon emissions reduction across fuel types.

Phase 1: Premise-Level Analysis

Phase 1 Methods

The subject evaluation began with a Qualtrics-based mixed-mode survey of approximately 4,500 customers that participated in the programs through 2018. The survey collected data on installed heat pump characteristics and use patterns, customer demographics, utility account information, and customer willingness to participate in future phases of the study. The survey responses provided an initial understanding of heat pump usage patterns, displacement of heating and cooling loads served by preexisting HVAC equipment, and occupancy fluctuations or other major non-routine events that occurred in the years since installation.² Bolstered by a \$15 incentive, the survey garnered 775 complete responses (17% success rate).

Phase 1 of the evaluation involved site-specific, weather-normalized analysis of pre- and post-installation monthly consumption data³ among the 434 survey respondents who provided viable account information. After collecting authorization from the respondents, evaluators obtained consumption data from 8 utilities and 45 suppliers of delivered fuels. Delivered fuel data collection is historically difficult. With 43% of the evaluation sample using heating oil or propane, delivered fuels represented an important segment for data collection. The evaluation team was able to achieve a 55% success rate in delivered fuel data collection by calling before

¹ Electric MMBtu at site does not account for generation or transmission losses.

² Since the evaluated installations occurred in 2018, the survey asked customers to disregard anomalous operation as a result of COVID quarantines in the post-installation period.

³ Advanced Metering Infrastructure (AMI) data was generally not available for participating customers in 2018.

the heating season intensified, batching calls (one dealer call for all authorizing customers), and using experienced callers instead of automated systems.

Next, we further screened the electric consumption data to ensure sufficient pre- and post-installation data (34 removed) and absence of solar photovoltaic systems, which can skew the pre/post analysis of heat pump impacts (78 removed). Anticipating sensitivity in the correlation of electric consumption with degree-day data, the team next developed three analysis filtering scenarios to present a range of evaluated results: mild, moderate, and strict. Each scenario incorporated increasing levels of strictness regarding data cleanliness, statistical significance, and weather dependency. For example, the team used lower R^2 thresholds than are sometimes used for residential HVAC analysis for the moderate and mild scenarios. This is due to observed heat pump use patterns. From our review of survey responses, we hypothesized that enough units were used in the fashion of appliances—with unit operation driven by occupant behavior almost as much as by outside air temperature—that excluding those without strong weather correlation would have likely biased the analysis.

The accounts passing the three screening scenarios were weighted to best represent the total population of program participants. We weighted results based on two stratification variables: equipment type (ASHP ducted, DMSHP, and GSHP) and climate zone (ICC 2011). Phase 1 billing analysis concluded with the quantification of program-wide savings, differentiated by the stratification variables, and by other variables of interest from the survey or tracking database.

Phase 1 Results

Figure 1 illustrates the achieved savings distinguished by the three screening thresholds. Using the moderate screening threshold, the billing analysis showed that ASHPs and GSHPs achieved per-project savings of approximately 15 and 46 MMBtu per year, respectively, across all fuels including electricity at site.

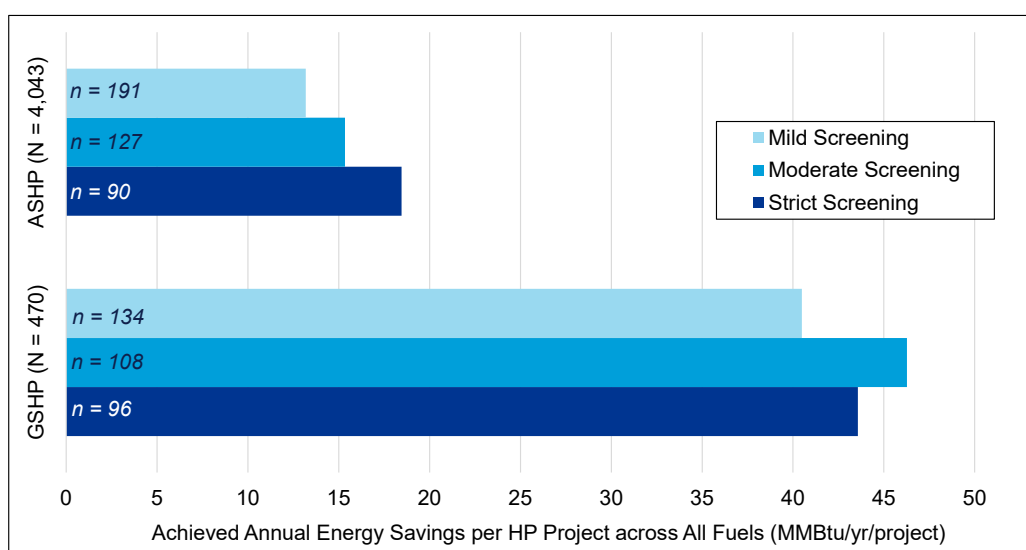


Figure 1. ASHP and GSHP Billing Analysis Results by Electric Screening Threshold

The authors compared billing analysis results among different variables that could influence heat pump savings, as self-reported by participants in the Phase 1 survey or determined

from program tracking databases. Evaluators identified three notable trends that affected achieved impacts in Phase 1:

- **Use of preexisting heating systems.** After reporting that their heat pump is used for at least some heating, respondents were asked how much their pre-existing heating system is still used. In cases where the customers continue to use their pre-existing heating system frequently, the resulting savings are significantly lower. Figure 2 illustrates that heat pumps used as the primary heating equipment realize over twice as much savings.

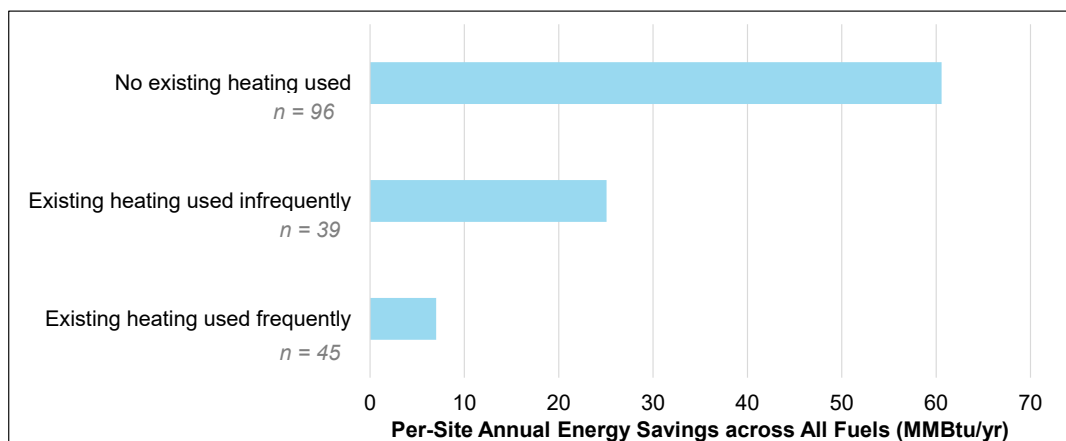


Figure 2. Billing Analysis Results as a Function of Preexisting Heating System Use (ASHPs and GSHPs Combined)

- **Location.** Urban customers realized lower savings per project than rural customers. We hypothesized that: 1) urban customers were more likely to install DMSHPs that conditioned a limited portion of the property's square footage, and 2) DMSHPs in multifamily settings were more likely to be exclusively used for cooling as a replacement for window air conditioners.
- **Project classification.** Installations at existing buildings outperformed those in newly constructed buildings. To define baseline consumption for new construction installations, we adjusted post-installation consumption data to reflect code-compliant systems self-reported by the customer as the preferred alternatives in the absence of the program. In some cases, heat pumps defined the baseline scenario, significantly reducing MMBtu savings as compared with fossil fuel alternatives.

Phase 2: Equipment-Level M&V

Evaluation Phase 2 involved M&V at a subset of the 588 survey respondents that expressed interest in hosting a site visit at their home or business. Evaluators visited 137 representative sites to confirm and refine survey response information, physically inspect the installed heat pumps, and deploy a combination of communicating and non-communicating meters. On-site data collection procedures varied depending on the installed HVAC equipment types and the selected level of metering rigor as described below.

Core M&V

Field evaluators executed core M&V procedures at 125 of the 137 sites in the evaluation sample. The core procedures represent industry-standard approaches for heat pump evaluation with the addition of advanced communicating devices, which operated using a data platform provided by a software-as-a-service (SaaS) contractor. The remote monitoring technology is designed to continuously transmit metered data to a secure, cloud-based repository via cellular network. Remotely monitored points included: compressor circuit amperage, outdoor fan circuit amperage, groundwater pump circuit amperage (for GSHPs), and supply air stream temperature and relative humidity. Field engineers supplemented the communicating meters with non-communicating meters on supply/return groundwater loop piping (for GSHPs) and various spot measurements of voltage, amperage, and real power at different heat pump operating modes and speeds.

Intensive M&V

For 12 sites evaluators conducted higher-rigor M&V with the objective of collecting real-world performance data to inform the 125 core sites and to reveal potential areas of focus in future studies. The intensive metering sites were selected to represent at least one of each major type of heat pump (e.g., ducted ASHP, DMSHP, GSHP) in each of the region's climate zones. The intensive approach supplemented the core protocol with long-term metering of true power on each of the relevant power circuits identified above, long-term metering of temperature and relative humidity before and after the heat pumps' coils, and spot measurements of delivered airflow at different modes/speeds using a flow hood. The supplementary intensive meters were non-communicating devices. Intensive site data allowed evaluators to compare at each metering interval the installed heat pump system's power draw with its delivered heating Btu. As heating efficiency increases (and cooling efficiency degrades) with increasing outside air temperatures, evaluators created curves that characterize the weather-dependence of each high-rigor site's heat pump performance.

Site-Specific Analysis Methods

Figure 3 illustrates the methodology for impact analysis of M&V data. Analysts processed metered data to characterize the heat pump's heating/cooling loads and operational patterns. Heat pump operation data was correlated with outside air temperatures over the metering period to determine the weather effects on heat pump operation. These correlations were extrapolated over a full year using typical weather data for the most proximate weather station. To quantify baseline energy usage, in the absence of non-routine events such as changes in occupancy, we assigned the heat pump's annualized heating output to the preexisting, code-compliant, or other alternative system. The evaluation team quantified annual energy impacts for each metered heat pump system by comparing the heating and cooling loads and associated performance efficiencies between baseline and as-built conditions. Final, site-specific impact results include savings or penalties by fuel: electricity, natural gas, and delivered fuels, as applicable. To develop program-wide savings, site-specific results were expanded to the population of participants using the M&V sampling weights defined by equipment type and climate zone strata.

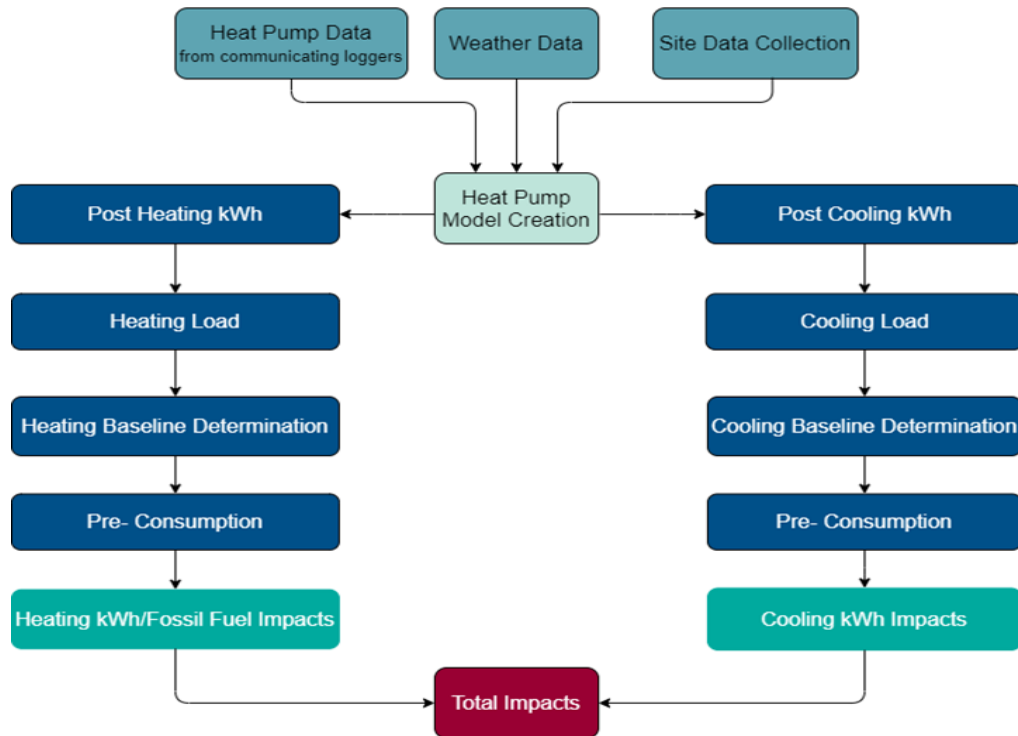


Figure 3. Impact Analysis Methodology for Phase 2 M&V Data

M&V Results

Figure 4 compares M&V results with billing analysis results (moderate screening).

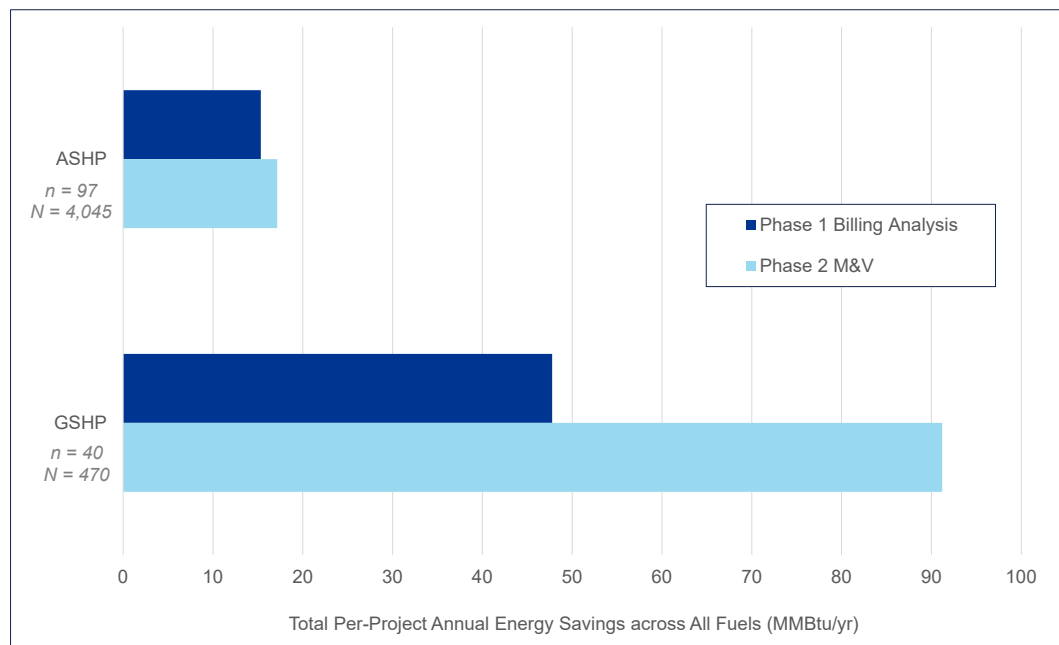


Figure 4. Comparison of Phase 1 and Phase 2 Achieved MMBtu Savings per Project

For ASHPs Phase 2 per-project savings closely resembled that of Phase 1, indicating the reasonableness of preliminary at-the-meter analysis to predict ASHP impacts. For GSHPs, on the other hand, we found significantly higher per-project savings in Phase 2 as compared with Phase 1. We attribute this difference to changes in baseline treatment. Phase 1 analysis was supported by customer survey responses on preexisting equipment and heat pump usage characteristics; Phase 2 involved more comprehensive, in-person interviews with customers on what they would have installed absent their participation in the program. For ASHPs we did not observe significant changes between survey responses and in-person data collection. But for GSHPs, we gathered more extensive information on feasibility and customer preferences of alternative systems (e.g., natural gas connectivity); this additional information resulted in baseline refinements for 16 of the 40 sampled GSHP sites that resulted in a higher prevalence of fossil fuel-fired systems, significantly increasing Phase 2 savings.

Similar to Phase 1, the evaluation team investigated results by variables of interest to explore opportunities for customer targeting and savings optimization. Figure 5 illustrates Phase 2 per-project results as a function of customer-reported responses to the question “Do you still use your preexisting heating system(s) to supplement the heat pumps?” Customers categorized under “no” include those who decommissioned the preexisting fossil-fuel heating equipment as well as electric heat customers that fully transitioned from the preexisting resistance system to heat pumps.

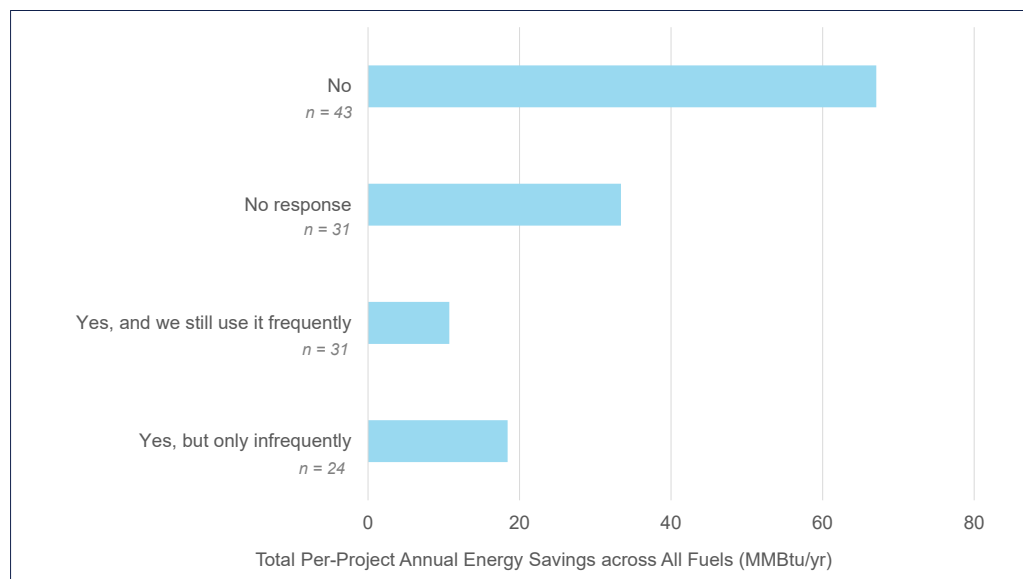


Figure 5. Phase 2 Results by Customer-Reported Use of Preexisting Heating System(s) (ASHP and GSHP combined)

Phase 2 showed similar variation as Phase 1, demonstrating that retention of preexisting heating systems was a strong predictor of achieved savings. Customers that de-commissioned the old systems achieved over twice the savings as customers that retained the systems. Usage frequency of legacy systems did not appear to be a strong indicator of performance.

Impacts by Fuel

Heat pumps are promising decarbonization measures due to their ability to efficiently offset fossil fuels. Quantification of fuel-specific impacts is therefore crucial to measuring the success of heat pump installations. Figures 6 and 7 illustrates the fuel-specific impacts, including beneficial electrification, for ASHPs and GSHPs, respectively. Beneficial electrification, as illustrated by the leftmost bar, indicates the new site-level electric heating load introduced by the rebated heat pumps as a result of displacing fossil fuels. The figure illustrates electric heating savings (striped orange bar), for conversions from electric resistance or less-efficient HP alternatives, as well as electric cooling savings (striped blue bar).

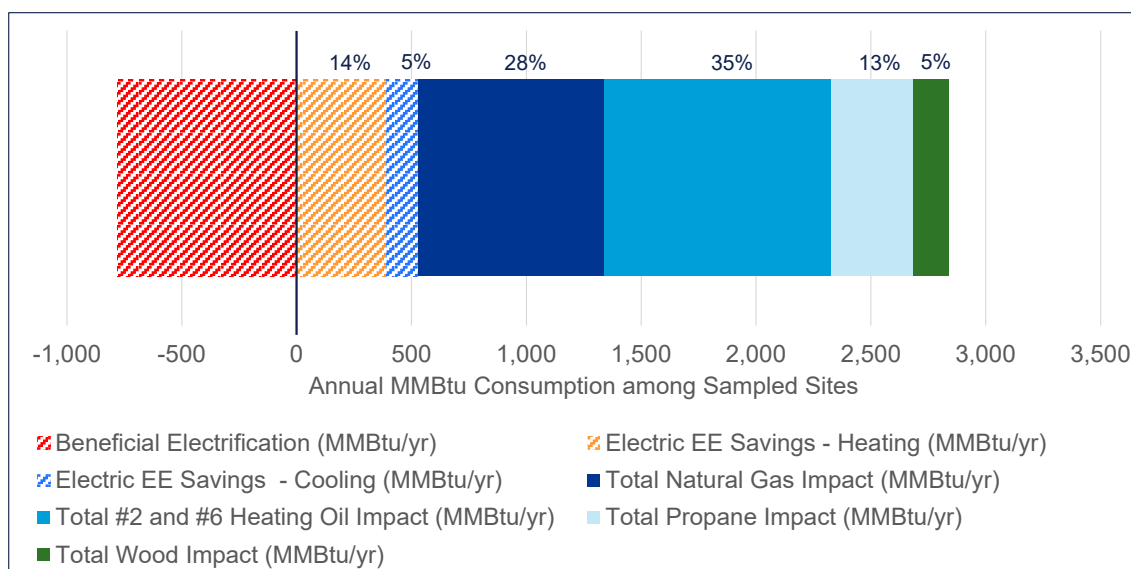


Figure 6. Phase 2 Achieved ASHP Heating Impacts by Fuel, Converted to Site MMBtu (Sampled Projects Only)

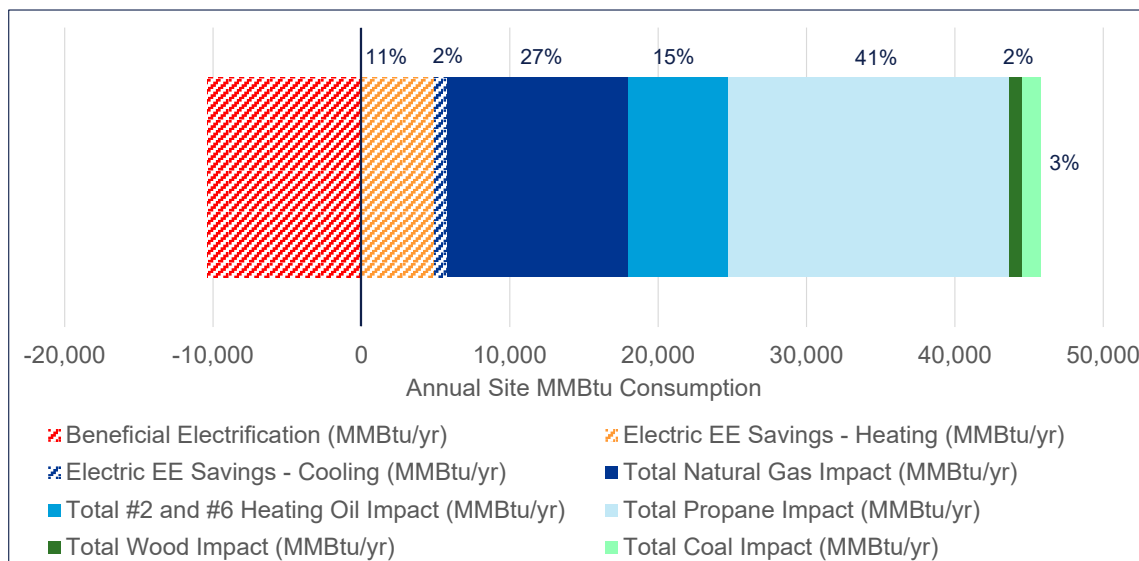


Figure 7. Phase 2 Achieved GSHP Heating Impacts by Fuel, Converted to Site MMBtu (Full GSHP Population)

For both ASHPs and GSHPs, evaluators determined significant savings shares for natural gas, fuel oils, and propane. Relative to heating savings among electric and fossil fuel alternatives, cooling savings accounted for only 5% and 2% of total achieved MMBtu savings for ASHPs and GSHPs, respectively. The evaluated heat pump projects occurred in the Northeast, where annual heating degree days far exceed cooling degree days. Additionally, the heating season offers far more savings potential when converting from fossil fuels—for every MMBtu (at site) of new electric load, the installed heat pumps saved approximately 3 MMBtu across all fuel sources.

Predicting Heat Pump Savings

Phase 1 billing analysis provided an initial assessment of at-the-grid impacts but limited the evaluators' ability to pinpoint reasons for unexpected heat pump performance. Phase 2 equipment-level M&V allowed evaluators to investigate heat pump operation more deeply and identify opportunities to right-size heat pump savings claims moving forward. The authors have identified three best practices for PAs when predicting heat pump savings:

- **Characterizing baselines** – As shown by the significant GSHP savings differences between Phase 1 and Phase 2, baseline considerations can drastically affect the achieved savings. We recommend that PAs consider site-specific constraints and customer preferences to inform the baseline fuel and system type for heat pump installations in new construction or end-of-life scenarios. Educating and monitoring a contractor base to credibly collect and track this information will be crucial to this recommendation's success. But accurately tracking presumed displaced fuel type is particularly critical for programs with electrification and carbon reduction goals.
- **Identifying displacements vs. replacements** – Figures 2 and 5 illustrate the impact of the preexisting heating system use on achieved heat pump savings. PAs should differentiate between partial displacements and whole-home replacements in savings estimation and incentive offerings. Higher incentives for both customers and contractors could stimulate more whole-home installations and achieve deeper savings.
- **Estimating heat pump output** – Phase 1 and Phase 2 illustrated that ASHPs generally did not operate as whole-home heating systems, as expected with the prevalence of DMSHPs. In Figure 8 evaluators compare the annual heating output of sampled ASHPs (vertical bars) with an average, whole-home, annual heating load value referenced from a Northeast Energy Efficiency Partnerships study on heat pumps (NEEP 2017).

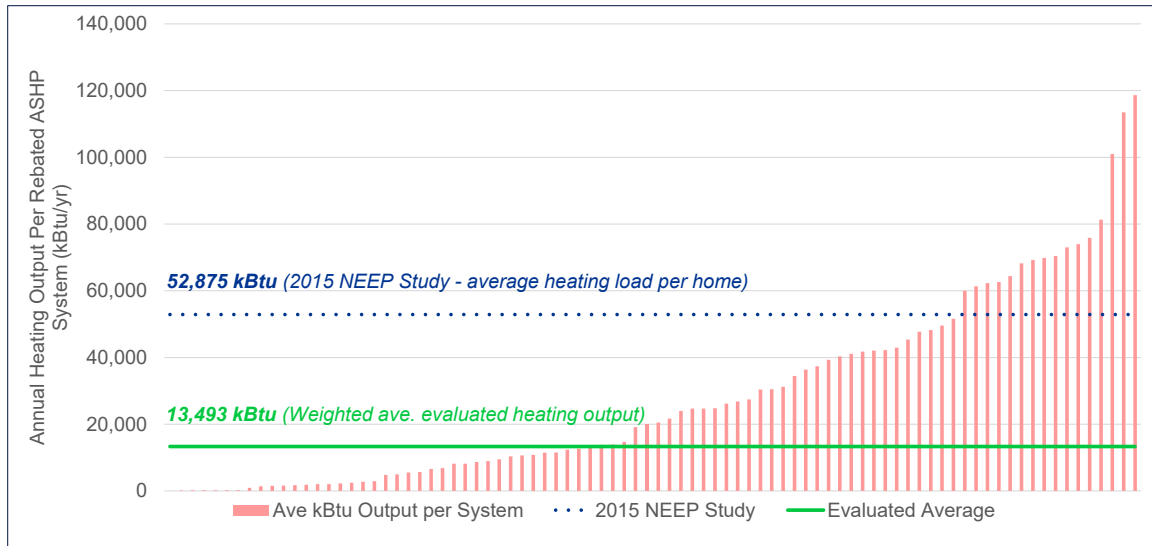


Figure 8. Comparison of Evaluated and Whole-Home Heating Output among Sampled ASHPs

On average, evaluated ASHPs delivered about a quarter of a typical home’s annualized heating load. After comparing equipment-level heating outputs with rated heating capacities, we determined an average equivalent full-load heating hours (EFLHH) value of 573 for ASHPs. Less frequent heating operation led to reduced opportunity for savings as compared with more carbon-intensive systems. The body of heat pump evaluation research is rapidly emerging, but other DMSHP studies in the Northeast have shown similar findings of approximately 450 EFLHH (MA/RI 2016). Given the prevalence of DMSHP partial-displacement installations, total installed capacity provides a more realistic starting point for “best case” savings, rather than a whole-home heating output value.

Billing Analysis vs. M&V

The evaluation upon which this paper is based employed two evaluation techniques to assess heat pump impacts, providing an opportunity for methodological comparisons. Both site-level billing analysis and equipment-level M&V provided useful results and offered comparative advantages. This section addresses how to choose between the two when seeking to independently quantify heat pump impacts. While there is no single best approach, answers to a series of questions regarding budget, schedule, and the program administrator (PA) objectives can help to make an informed choice. All of this discussion presumes the program collects sufficient information to identify and contact residents that received the heat pumps; this presumption may not be true for upstream or midstream programs. Figure 9 summarizes the advantages and disadvantages of the two methods, including three levels of rigor for premise-level analysis.

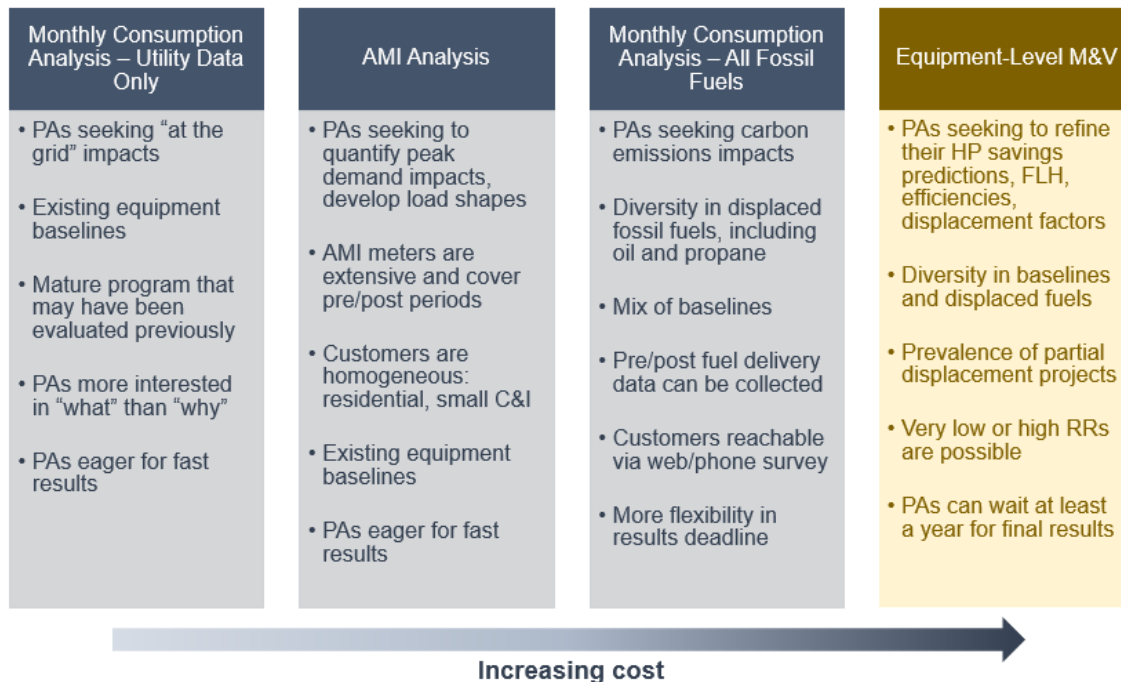


Figure 9. Methodological Comparison of Premise-Level Analysis and Equipment-Level M&V

Starting from the left, the most economical and fastest method to quantify heat pump impacts is **premise-level analysis of monthly consumption data** covering the pre- and post-installation periods. Consumption data **exclusively available from utilities**—i.e., electric and natural gas—streamlines the data collection process and reduces cost. A prevalence of early replacement installations, for which the pre-existing equipment defines the baseline condition, makes the analysis even more economical and expeditious. However, this lower-rigor approach is unlikely to reveal the key drivers to differences in evaluation results as compared with program predictions. For this reason, this approach is best suited for mature, high-volume programs that may have already been evaluated to refresh results with a new tranche of projects.

PAs may be interested in more than just energy impacts. As **advanced metering infrastructure** (AMI) propagates, its 15-minute data allows for more rigorous and dynamic analysis of heat pump impacts. AMI data lets evaluators develop peak coincidence factors and hourly load shapes that can inform not only evaluation results but also utility forecasts and rate case adjustments. A number of criteria must be met for AMI candidacy: the meters cover sufficient pre- and post-installation periods, the project baselines are generally informed by pre-existing equipment, and the AMI profiles of the studied customers are generally homogeneous. A prevalence of large C&I participants, for example, limits the use of AMI data.

The third method, **site-specific analysis of premise-level energy consumption across all fuels**, closely resembles this evaluation’s Phase 1. As demonstrated with ASHPs, the approach provided reasonable, preliminary savings results, even with fuel switching. It was not presumed in advance to be as effective for partial displacement heating conditions—a common scenario for DMSHPs—but proved to be. Effective premise-level consumption analysis required in-depth surveying, a relatively large sample, and collecting and cleaning consumption data for all relevant fuels including unregulated sources such as oil and propane. Phase 1 also required site-specific adjustments to account for new construction or normal replacement baseline conditions. For these reasons, it is a costlier and lengthier alternative to the prior two methods

but is the most economical option to provide comprehensive, fuel-by-fuel energy and carbon impacts.

The fourth method, **equipment-level M&V**, is the most accurate method for individual site estimation and closely resembled this evaluation's Phase 2. It has more explanatory power than other methods. This is particularly important when the evaluated savings vary dramatically, as they did in the subject evaluation. Metering heat pumps directly shows not only how many full load hours they operate (though, sometimes premise-level analysis can do this as well) but also how frequently the heat pumps run, how loaded they are when they do operate, and—when other heating sources are also metered simultaneously—how often the heat pumps are primary or secondary heat sources. This information is critical for PAs seeking to improve savings estimation moving forward. Regarding calendar time to results, this option requires more time than premise-level analysis because the typical year of post-installation data collection cannot start immediately upon unit installation. The possibility of near-real-time data collection reduces but does not eliminate this drawback compared to using utility data.

Conclusion

The subject heat pump evaluation demonstrated the benefits and shortcomings of two evaluation approaches. Phase 1 analysis of premise-level consumption data provided preliminary ASHP and GSHP savings estimates of 15 and 46 MMBtu savings, respectively, per year per project across all fuels, including electricity at site as well as delivered fuels. Phase 2 M&V results closely resembled Phase 1 for ASHPs—17 MMBtu savings per year per project—confirming the reasonableness of preliminary at-the-meter analysis of ASHP installations. However, GSHP savings in Phase 2 nearly doubled the preliminary estimate in Phase 1 with 91 MMBtu savings per year per project. This significant increase was primarily due to changes in baseline treatment as a result of in-person data collection by evaluators. Evaluators revised the baseline for 40% of sampled GSHP projects after determining the most appropriate alternative system type based on feasibility and customer preferences. These baseline revisions largely resulted in a shift toward fossil fuel-fired systems, increasing MMBtu savings at site.

While significantly less expensive and more expeditious than Phase 2, Phase 1 was limited in its ability to characterize nuanced baseline scenarios and to pinpoint why evaluated savings deviated from expectations. Phase 2 equipment-level M&V data allowed evaluators to comprehensively examine savings variation and recommend best practices for estimating scalable, accurate savings for heat pump installations that are expected to grow by orders of magnitude in the next decade. Such best practices include collection and tracking of credible site-specific data on preexisting systems and alternative choices, application of customized baselines established by that information, and use of rated equipment capacities to determine reasonable savings values in partial displacement scenarios. Phase 2, however, required significant budget and time commitments, though remote-communicating capability allowed the team to continuously analyze the data among different seasons.

The authors have summarized the advantages and disadvantages of premise-level analysis and M&V methods depending on program complexity, budget, data availability, and time constraints. While these best practices are presented in the context of a heat pump evaluation, their principles apply to any assessments of programs or technologies with varying levels of data availability.

References

- ICC (International Code Council). 2011. *2012 International Energy Conservation Code*, Section C301.1. <https://codes.iccsafe.org/content/IECC2012>
- MA/RI (Massachusetts and Rhode Island Electric and Gas Program Administrators). 2016. “Ductless Mini-Split Heat Pump Impact Evaluation.” <http://www.ripuc.ri.gov/eventsactions/docket/4755-TRM-DMSHP%20Evaluation%20Report%2012-30-2016.pdf>
- NEEP (Northeast Energy Efficiency Partnerships). 2017. “Northeast/Mid-Atlantic Air-Source Heat Pump Market Strategies Report 2016 Update.” https://neep.org/sites/default/files/NEEP_ASHP_2016MTStrategy_Report_FINAL.pdf