Measuring up -- how does my baseline compare?

Riley Hastings, Eversource Energy, Westwood, MA
Justin Spencer, Navigant Consulting, Boulder, CO
Terese Decker, Navigant Consultant, Boulder, CO

ABSTRACT

A group of Massachusetts Program Administrators (PAs), evaluators, and other stakeholders are undertaking a statewide study to understand more about how residential buildings and their occupants consume energy. Through this study, we will learn about relevant opportunities for reducing overall consumption and consumption during peak periods. The PAs have the ambitious goal of quantifying the load shape for all major current and future electric end uses in addition to obtaining their current saturations, in order to provide the basis for future program planning in the changing world of distributed energy resources. Because of the complexity and expense of conducting this study, the team is conducting the work in two phases. Phase I (May 2016 – December 2016) tests and verifies the data collection and analysis approach on about 5% of the overall sample of sites before proceeding with Phase II (January 2017 – Spring 2018), a full-scale implementation of the study. This paper discusses how a phased approach can be used in a large scale study such as this one to reduce the overall risk and cost associated with the study and highlights the changes made to the study design based on the findings from Phase I. By testing out cutting edge approaches before deploying them on the whole study, the PAs are able to "try before we buy," ensuring that the Massachusetts stakeholder group gets what it needs from this major three million dollar undertaking.

Introduction

The objective of this statewide Massachusetts baseline study is to understand how residential buildings and their occupants consume energy. The PAs have the ambitious goal of quantifying the load shape for all major current and future electric end uses in addition to obtaining their current saturations, in order to provide the basis for future program planning in the changing world of distributed energy resources.

Because of the complexity and expense of conducting this study, a two-phase process has been employed for conducting the study. In the first phase, the evaluation team designed research experiments, developed data collection tools and protocols, and then tested them by conducting the full data collection and analysis on about 5% of the overall sample. In this second phase, the evaluation team revised all data collection and analysis protocols and tools to better meet the objectives of the study and conduct the full rollout of all data collection and analysis. The results of the first phase of the study have helped to inform the experimental design of the second phase.

More specifically, the purpose of Phase I was to:

- Develop and rigorously test onsite data collection tools and protocols.
- Determine the predictive power of Non-Intrusive Load Monitoring (NILM) for each end use being tested and refine the circumstances under which NILM will prove useful for Phase II.
- Determine the accuracy of individual saturation survey questions and identify new questions or responses that should be added in Phase II.
- Verify the quality of data gathered using different protocols (surveys, NILM, and on-site metering).

Based on the totality of findings in Phase I, the team made significant revisions which include:

- Moving all surveys to online
- Eliminating NILM from the onsite
- Increasing the number of on-site intensive end-use metering sites
- Adding lighting data collection to the onsite validation protocols
- Making survey changes

Background

The Massachusetts PAs decided to undertake this study for several reasons. The PAs need this data for potential studies that the MA Department of Public Utilities (DPU) ordered the PAs to undertake in support of their 2019 to 2021 three-year plan goals. Similarly, the PAs also were in need of home characteristics and appliance saturation data. The last time Massachusetts saturation study was completed in 2009. For example, one of the interesting takeaways that we have already learned based on draft survey results data is that the percentage of homes with oil heat has declined from 36 percent overall in 2008 to 19 percent in 2017. This is a significant part of the benefits achieved by the PAs in their Home Energy Services program and explains why the number of homes with oil heat being weatherized has been declining at such a steady and fast rate.[1]

As discussed above, the PAs have the ambitious goal of quantifying the load shape for all major current and future electric end uses in addition to obtaining their current saturations. Currently the residential load shapes used by the Massachusetts PAs are based on some primary research studies but many are estimated based on secondary research. The PAs are in the process of launching demand pilots and anticipate that demand will be a greater focus of energy efficiency programs going forward. Having accurate load shapes is an important component of accurately estimating those demand savings.

Results

This section presents the most significant lessons learned from Phase I, and discusses how the results have impacted the full scale roll out of the research for Phase II.

Surveying Method and Sample Size

After analyzing the results from Phase I, Navigant found that the online survey was a better representation of the population than the phone or paper survey, based on a comparison of phone and survey participant demographics to American Community Survey demographics for the Phase I test region. In addition, administering the survey via paper or phone is significantly more expensive than offering it online. Since so few people even asked to do a paper or phone survey Navigant did not advertise these as options in Phase II. Navigant only administered phone surveys upon request in Phase II if people did not have access to a computer or internet.

In Phase I Navigant included a phone number for a Navigant employee on the postcards for respondent to call if they had questions on the survey, but in Phase II they changed it to an e-mail address that was accessible to multiple Navigant employees. Having people respond to an e-mail address instead of a phone number proved to be a much more efficient and effective way for addressing customer questions in Phase II. There were a few people who chose to write letters to the survey house because they did not have a computer to send an e-mail. In those instances, Navigant followed up with the customer via a phone call.

[1] The Massachusetts PAs can claim benefits for weatherizing homes with delivered fuels.

The original plan was to survey 10,000 customers, at the same time recruiting for an on-site study with a goal of 1000 NILM households and 140 full end use metering households. However, once the team learned that we were able to recruit people more easily for onsite visits then we originally thought and we were able to achieve subsegment precision targets with 6,300 online surveys and 478 onsites.

Eliminating NILM for Onsite

Based on the results and findings regarding the cost of data collection for whole house and end use metering approaches, Navigant's recommendation was that the Phase II Massachusetts Residential Baseline not depend on disaggregation of whole house data. The reasons for this recommendation were the relative cost of whole house monitoring and whole house with end metering data collection, the quality of the data collected via low cost whole house monitors, and the inconsistent accuracy of the estimated end use load curves based on disaggregation techniques such as NILM.

In general, Navigant found evidence that there are opportunities to leverage disaggregation (NILM) for the estimation of several end uses. However, there were limitations on the ability to disaggregate many of the hardwired loads, which are more challenging to meter directly. This is where the combined disaggregation and sub-metering approach would have the greatest value, but it did not appear as though this approach is sufficiently developed as yet to achieve this aim. The NILM work is explained in greater detail in another IEPEC paper (Elszasz, 2017).

The original plan included 1,000 whole house energy monitored (NILM) sites. Throughout the Phase I data collection, Navigant found that the average per site cost of end use metering was less than anticipated. In contrast, the average per site cost of whole house metering via home energy monitors was greater than anticipated. This difference was primarily due to two factors:

- Fewer hours required per site for the set-up and collection of end use metering data, and
- Additional time necessary to ensure that the whole house energy monitors stayed online and provided continuous, usable data

Depending on the specific objectives of the study, it may be reasonable to proceed with the cheaper and less accurate metering option (Decker, 2017). However, in this case the team decided the risk of sacrificing accuracy was too great to justify taking the less expensive route. Table 1 shows a comparison and contextual information about the two data collection methods tested in Phase I.

Table 1. Data Collection Methods Comparison

	Whole Home Energy Monitor	Full End Use Metering
Equipment Cost*	$410 / site	$1,250 / site
Installation Qualifications	None	Electrician
Installation time	0.5 hours / site for one person	3 hours / site for two people
Maintenance time	2.5 hours / site	0.5 hours / site
Data Frequency	32-second	1-minute
Data Quality	Poor	Excellent
Analysis Required	Third party load disaggregation	In-house
Within-home Communication	Home internet bridge	Powerline carrier

*Equipment cost includes shipping, installation training, data access, and technical support

Increasing number of full end use metering sites

Since the team decided to eliminate homes with whole home energy monitors, that cost was put towards increasing the number of sites with full end use metering from 140 sites to 478. To further decrease the cost of metering and increase the usefulness of the study, the team chose to optimize the sample by splitting the sample into 300 core sites and 178 oversample sites. The core sites in the sample have all possible data collected at each site and will be used to draw overall conclusions about the state's appliance saturations and energy usage. Oversample sites are only focused on a single end use, and are only used to increase the precision of the energy usage estimates for that single end use.

At each oversample site visit, the field technician will gather data about the end use of interest and meter only the end use of interest. At sites where the oversampled end use requires metering at the electrical panel, the field technician will also meter whole home energy consumption. Table 2 shows the planned level of rigor by end use where the least rigorous is saturation and most rigorous is loadshape which includes a saturation, characterization and loadshape

Table 2: Planned Level of Rigor by End Use

Category	Loadshape	Characterization	Saturation
Heating and Cooling	Central AC / heat pump Room AC Other electric heat Other fuel furnace fan Space heaters / plug in fireplaces Boiler circulator pump	Ductless AC / heat pump Ground source heat pump Whole house fan Thermostat HRV/ERV or other central mechanical Gas furnace Gas boiler Other fuel furnace Other fuel boiler Gas fireplace Other fuel other heat	Ceiling fan
Kitchen	Dishwasher Refrigerator Second refrigerator Freezer	Coffee maker Gas oven Electric oven Other fuel oven Kitchen fan / range hood	Electric range Microwave Gas range Gas grill Other fuel range Instant hot water dispenser Toaster / toaster oven Blender / Juicer
Hot Water	Electric water heater Heat pump water heater	Electric tankless water heater Hot water recirculator pump Gas water heater Solar water heater Other fuel water heater	Showerheads Faucets
Laundry	Electric dryer Electric washer	Gas dryer	
Misc.	Pool pumps Well pumps Sump pumps Dehumidifiers Booster pumps	EV charger Electric hot tub Gas hot tub Freeze prevention heat Electric pool heater	Smart phones Tablets Air cleaner Gas lighting

Category	Loadshape	Characterization	Saturation
	Ejector pumps	Gas pool heater Lighting Other garage chargers Humidifiers Home entertainment equipment Televisions Home office equipment Water beds Engine block heater Fish tanks Golf cart / large battery charger	

In addition to testing out the different data collection methodologies considered, Phase I also helped eliminate uncertainty around coefficient of variation (CV), especially during peak times and for end uses that had no other good source of information. The CV assumption is frequently the most difficult assumption to derive from another source and it has a significant impact on precision.[2] Other sources of CVs did not exist or included only CVs on total energy use instead of peak. Given the focus of this study on peak demand, providing some kind of indicator of the validity of CV assumptions was very helpful to the development of the phase II sample. Table 3 below shows the estimated frequency and CVs used to develop the sample sizes. It should be noted that some of the CVs for peak are much higher than CVs typically used for energy.

Table 3: Estimated frequency and CV used to develop sample sizes

End Use Category	Target End Use	Frequency of Occurrence in MA Households	Target Precision	CV - Energy	CV - Utility Peak	CV - ISO Peak	Oversample Driver
Heating and Cooling	Central AC / Heat Pump	29.0%	15.0%	0.56	0.60	0.56	CV - Utility Peak
	Room AC	64.0%	30.0%	1.33	1.70	2.36	CV - Utility Peak
	Ground Source Heat Pump	1.0%	100.0%	0.56	0.60	0.56	CV - Utility Peak
	Other Electric Heat	5.0%	20.0%	0.50	2.00	1.88	CV - Energy
	Space Heaters / Plug-in Fireplaces	15.0%	15.0%	1.00	2.45	2.45	CV - Energy
	Boiler Circulator Pump	51.0%	15.0%	0.50	0.77	0.43	CV - Energy
	Other Fuel Furnace Fan	53.0%	10.0%	0.50	0.15	0.13	CV - Energy
Kitchen Appliances	Dishwasher	71.0%	30.0%	0.79	1.93	2.66	CV - Utility Peak
	Freezer	13.0%	20.0%	0.42	0.32	0.38	CV - Utility Peak
	Refrigerator	100.0%	20.0%	0.31	0.32	0.28	CV - Utility Peak

[2] For a given confidence level and sample size, the resulting precision goes up proportionally with CV. E.g. a sample designed to achieve 90/20 with a 0.5 CV assumption will achieve 90/40 if the CV is actually 1.0.

End Use Category	Target End Use	Frequency of Occurrence in MA Households	Target Precision	CV - Energy	CV - Utility Peak	CV - ISO Peak	Oversample Driver
	Second Refrigerator	28.0%	20.0%	0.59	0.54	0.53	CV - Utility Peak
DHW	Hot Water Heater	15.0%	15.0%	0.42	0.65	0.93	CV - ISO Peak
	Tankless Hot Water Heater	0.6%	100.0%	0.42	0.65	0.93	CV - Utility Peak
	Heat Pump Water Heater	2.0%	20.0%	0.42	0.65	0.93	CV - Utility Peak
Laundry	Washer	86.0%	30.0%	0.68	2.17	2.47	CV - Utility Peak
	Electric Dryer	68.1%	20.0%	0.46	3.81	1.52	CV - ISO Peak
Miscellaneous	Dehumidifier	50.0%	15.0%	0.83	0.83	0.76	CV - Utility Peak
	Aquarium	2.0%	100.0%	0.50	0.50	0.50	CV - Utility Peak
	Golf Cart/Large Battery Charger	5.0%	100.0%	1.00	1.00	1.00	CV - Utility Peak
	Well Pump	12.0%	100.0%	1.72	1.84	1.62	CV - ISO Peak
	Sump Pump	30.0%	30.0%	1.44	1.48	1.46	CV - Energy
	Booster Pump	5.0%	100.0%	0.80	0.06	0.87	CV - Energy
	Pool Pump	10.0%	20.0%	0.72	0.75	0.79	CV - ISO Peak
	EV Charger	0.4%	100.0%	0.70	1.00	1.00	CV - Utility Peak

Table 4 below shows the resulting precisions, estimated samples achieved for each end use within the core group of 300, and the resulting number of oversample sites that were selected.

Table 4: Precisions by target end use

End use	Estimated Frequency	Targeted End Use Precision - Energy	Targeted End Use Precision – Utility Peak	Targeted End Use Precision – ISO-NE Peak	Oversample	Number of Core Sample Sites*	Number of Oversample Sites	Total Number of Sites
Central AC / Heat Pump	29.0%	10.1%	10.6%	9.9%	No	87	0	87
Room AC	64.0%	15.8%	20.3%	28.2%	No	192	0	192
Ground Source Heat Pump	1.0%	84.3%	100.6%	93.6%	No	3	0	3
Other Electric Heat	5.0%	20.5%	82.0%	77.2%	Yes	15	3	18
Space Heaters / Plug-in Fireplaces	15.0%	15.0%	36.8%	36.8%	Yes	45	77	122
Boiler Circulator Pump	51.0%	6.7%	10.3%	5.8%	No	153	0	153

End use	Estimated Frequency	Targeted End Use Precision - Energy	Targeted End Use Precision – Utility Peak	Targeted End Use Precision – ISO-NE Peak	Oversample	Number of Core Sample Sites*	Number of Oversample Sites	Total Number of Sites
Other Fuel Furnace Fan	53.0%	6.6%	1.9%	1.7%	No	159	0	159
Dishwasher	71.0%	9.0%	21.8%	30.1%	No	213	0	213
Freezer	13.0%	11.2%	8.7%	10.4%	No	39	0	39
Refrigerator	100.0%	3.0%	3.1%	2.7%	No	300	0	300
Second Refrigerator	28.0%	10.7%	9.9%	9.6%	No	84	0	84
Hot Water Heater	15.0%	6.7%	10.5%	15.0%	Yes	45	61	106
Tankless Hot Water Heater	0.6%	186.1%	289.8%	415.6%	No	2	0	2
Heat Pump Water Heater	2.0%	13.0%	20.3%	29.1%	Yes	6	24	30
Washer	86.0%	7.0%	22.3%	25.4%	No	258	0	258
Electric Dryer	68.1%	5.4%	44.1%	17.6%	No	204	0	204
Dehumidifier	50.0%	11.2%	11.3%	10.3%	No	150	0	150
Aquarium	2.0%	41.1%	41.1%	41.1%	No	6	0	6
Golf Cart/Large Battery Charger	5.0%	45.5%	45.5%	45.5%	No	15	0	15
Well Pump	12.0%	48.6%	51.8%	45.6%	No	36	0	36
Sump Pump	30.0%	25.3%	25.9%	25.6%	No	90	0	90
Booster Pump	5.0%	36.3%	2.9%	39.6%	No	15	0	15
Pool Pump	10.0%	18.2%	19.1%	20.2%	Yes	30	14	44
EV Charger	0.4%	N/A	N/A	N/A	No	1	0	1
Total						300	178	478

Based on the planned levels of rigor, Table 5 below shows the planned number of oversamples by end use. The ultimate number of oversamples will depend on the observed saturations in the population and resulting core sample.

Table 5. Oversample Scope by End Use

End Use	Number of Oversamples	Meter Type	Onsite Scope
Hardwired electric heat	3	Hardwired	Meter all hardwired electric heat onsite, collect characterization data about each, meter whole home energy consumption
Plug-in space heater/fireplace	77	Plug-in	Meter all plug-in space heaters onsite, collect characterization data about each

End Use	Number of Oversamples	Meter Type	Onsite Scope
Electric hot water heaters	61	Hardwired	Meter all electric hot water heaters onsite, collect characterization data about each, meter whole home energy consumption
Heat pump water heaters	24	Hardwired	Meter all hardwired electric heat onsite, collect characterization data about each, meter whole home energy consumption
Pool pumps	14	Some hardwired, some plug-in	Meter all pool pumps onsite, collect characterization data about each, meter whole home energy consumption if pool pump is hardwired in whole home panel

Adding Lighting Assessment

In Phase II, the team is also collecting lighting data at each site to support a comprehensive data set where the PAs can understand lighting saturations in the context of other home characteristics. While we do a lot of lighting research in the state where lighting saturations are collected and known, the team wanted this study to be comprehensive. We decided that it was not desirable to have the lighting assessment separate from rest of the home characteristics.

Survey Design Changes

Implementing the online survey in Phase I allowed Navigant to test different approaches and identify the best path forward for Phase II with respect to response rate, question effectiveness, survey time, and ease of use. The list below summarizes the key changes that Navigant made to the online survey in Phase II due to the lessons learned in Phase I.

1. ***Added photos of equipment types to eliminate customer confusion.*** In Phase II Navigant added pictures of heating, cooling, and water heating equipment to eliminate customer confusion on what type of equipment customers had in their home.
2. ***Adjusted the structure of questions to make them more user friendly.*** In Phase II Navigant adjusted the framing of questions to make them easier to understand in hopes of getting more accurate survey responses and reducing the customer's time required to take the survey.
3. ***Added questions that would be valuable for program design purposes based on stakeholder feedback.*** After the conclusion of Phase I Navigant collected feedback from stakeholders on which questions they would like to add that would be helpful for research and program design purposes. For example, in Phase II Navigant added questions on ejector pumps and booster pumps based on stakeholder feedback.
4. ***Removed questions that did not serve a research purpose based on stakeholder feedback.*** Navigant worked with stakeholders to remove questions that did not serve a research purpose, which helped reduce the time required for customers to take the survey.
5. ***Added clarification to questions that appeared to be causing customer confusion.*** After the conclusion of Phase I both Navigant and key stakeholders did a top to bottom review of the survey. Out of that review process, Navigant added clarification to a few questions to minimize customer confusion about what the question was asking.
6. ***Removed questions where many people dropped out of the survey after they were asked that question.*** For example, in Phase I many people dropped out of the survey when they were asked to enter their pin number. In Phase II Navigant added a picture of the postcard with the location of the pin number circled so less people would drop out of the survey. Also, in Phase I many people dropped out of

the survey after they were asked which hours of the day their home was occupied. In Phase II Navigant modified this question so that it was less intrusive. Instead of asking when the customer's home is occupied for each hour and day of the week Navigant asked customers more qualitatively when their homes are occupied on weekdays and on weekends.

7. ***Adjusted the combinations of survey modules that were asked to limit the average time in survey to under 30 minutes.*** After the conclusion of Phase I Navigant analyzed how long it took customers to fill out each module of the survey. Based on that analysis Navigant adjusted which survey modules were paired together. For example, in Phase I Navigant found that the kitchen appliances module and the miscellaneous non-plug loads module should not be paired together because they are time intensive modules.

8. ***Removed questions that were not resulting in useful responses.*** For example, in Phase I Navigant asked people who said they own a pool whether their pool has a filter and a pump and 100% of respondents said yes. As result, Navigant removed this question in Phase II.

9. ***Population targeting.*** The Phase I survey response rate was 6%, with ~3% response rates for both the first postcard and the reminder postcards. In addition, the response rate among renters was lower than expected, which was ultimately traced to an issue with the population frame. In the revised Phase II survey effort, the team used a new population frame based primarily on electric PA customer lists and continued to check demographics against ACS data. Based on the response rates, the team decided to try adding a third recruitment postcard. Unfortunately, in early Phase II results, the response rates were lower than expected and the third postcard was not very effective. Judging response rates from multiple postcards is not easy because of the long tails for responses – in some cases responses have come in up to 2 months after the last reminder postcards to early recruits. Because of the low response rates, the team added geographically targeted recruitment to fill in demographic shortages. The team looked at response rates for ACS PUMAs and the corresponding demographics of their responses. The team then targeted the last wave of surveys to the PUMAs that had both low response rates and high rates of the underrepresented demographics (low income, multi-family, and non-English speakers). For this final wave of postcards which were intended to true up the demographics, the team sent a single postcard to more residents rather than multiple postcards to fewer residents because the first postcard was found to be most effective.

10. ***Data collected on-site vs online.*** During the Phase I survey, the team also tested the capability of web survey respondents to answer certain questions by comparing with the data collected onsite. For Phase II, the team decided to only collect data in one place, either onsite or in the survey. The team collected size and efficiency characteristics onsite, while collecting square footage only the survey. This decision was optimized based on respondents' ability to answer questions dependably.

Conclusion

Based on the totality of findings in Phase I, the team revised all the data collection protocols and tools in order to ensure the seamless linking of all data points. These revisions included:

- Revising all data collection tools and protocols based on lessons learned in Phase I, including eliminating NILM from the onsite and analysis protocols and adding lighting data collection to the onsite validation protocols;
- Phrasing questions in the survey to ensure respondent understanding matches the evaluation team's intent;
- Reducing redundancy between online and onsite survey for customer-reported questions (such as house square footage) where there is no compelling reason to ask the same question twice;
- Removing questions from the survey that did not prove valuable in the analysis;

- Compiling open ended "other" responses into multiple choice answers for easier analysis; and
- Ensuring that the data linkages work seamlessly between all data collection methods.

The final overall methodology combines a series of nested data collection techniques to derive answers to the research questions. **Error! Reference source not found.**~~Figure 1~~ shows the different types of data collection the research team will employ by research question. In this approach, each level of low-cost data collection is trued up by a more rigorous level of data collection. The combination of low-cost data collection with high-rigor data collection will allow for reduction of both sampling bias and measurement error in the study results, while managing overall costs.

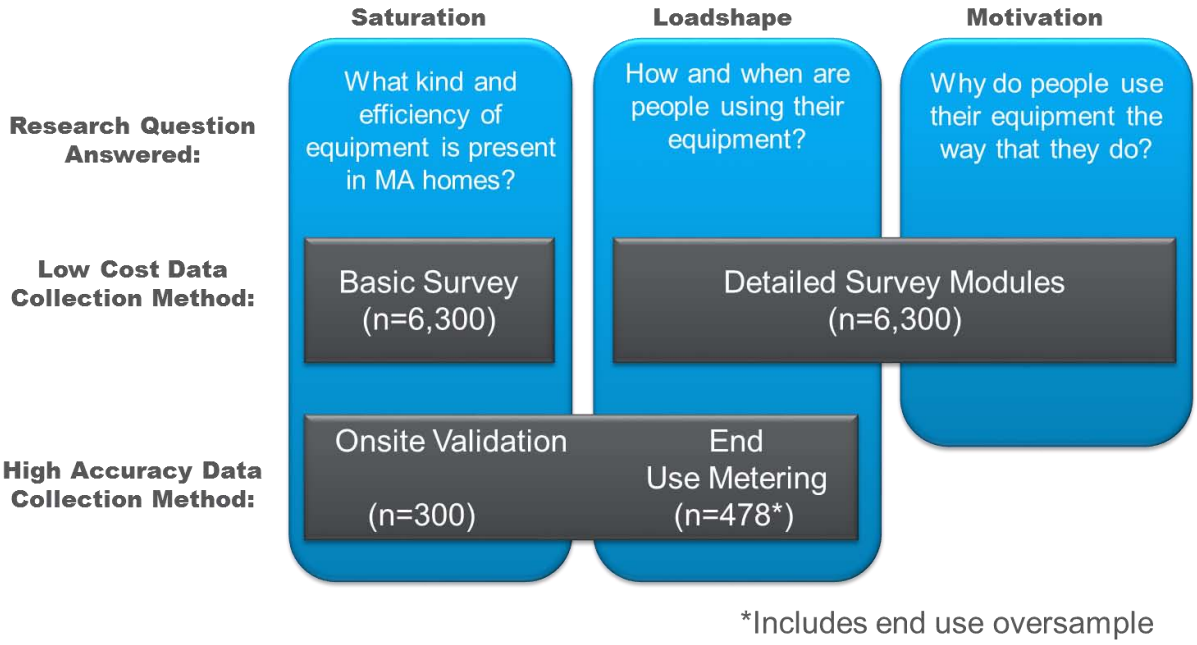

Figure 1. Nested Data Collection Approaches and Flow Diagram

The PAs envision being able to use this robust data for many purposes in the future. Most immediately, this data will be used to support the PAs potential studies being used to support setting goals in their 2019 to 2021 three-year plans.

Under the PAs Home Energy Services (HES) core initiative, enhanced incentives are provided to Moderate Income households, defined as 1-4 unit households with incomes between 60 and 80 percent of the state median income (SMI). The PAs are assessing the potential for expanding the offering to households from 81%-100% and 101%-120% of the SMI. This baseline study will serve as a sample source for a follow-up survey of potentially moderate income eligible customers, a group that would be hard to identify without this study, as well as be mined to provide an initial population size estimate and building characterization.

Some other possible uses of this data include: (1) forecasting and following trends such as customers converting from oil, (2) keeping track of saturations of emerging technologies such as electric vehicles, and (3) providing value to other areas of the company such as forecasting and transmission and distribution who are interested in load shape data but do not necessarily have access nearly 500 end use metered sites.

References

Opinion Dynamics 2009. “Massachusetts Residential Appliance Saturation Survey (RASS).” Prepared for the Massachusetts PAs.

Decker et al. 2017. “Duckhunt! Benefits and risks of load disaggregation and end use metering for determining end use loadshapes.” International Energy Program Evaluation Conference, Baltimore, MD. August 8-11.

Elszasz et al. 2017. “A Snapshot of NILM: Techniques and Tests of Non-Intrusive Load Monitoring for Load Shape Development.” International Energy Program Evaluation Conference, Baltimore, MD. August 8-11.