CONDUCTING A WEB-BASED, RESIDENTIAL APPLIANCE SATURATION SURVEY (RASS) ACROSS MULTIPLE ELECTRIC COOPERATIVES: METHODOLOGY, CHALLENGES, RESULTS, AND NEXT STEPS

Michael Witt, Justin Holtzman and Madhur Lamsal, DNV GL, Arlington, VA
Kyle Wood, Seminole Electric Cooperative Inc., Tampa, FL

ABSTRACT

DNV GL worked with Seminole Electric Cooperative Incorporated (Seminole) in 2016 to design and implement a residential appliance saturation survey (RASS) across each of their nine Distributor Member Cooperatives' service territories. The purpose of the study was to generate precise, statistically defensible and comparable saturation estimates and energy usage patterns among customers residing in the territories. Estimates associated with a variety of topics were developed that covered categories such as dwelling unit-level structural characteristics, space heating and cooling, electronics and appliances, energy conservation improvement activities and household demographics.

A RASS has several applications in the energy industry and is typically conducted to support long-term forecasting for asset management and capacity planning purposes, to further understand energy consumption characteristics of customers' end uses, and to inform energy efficiency and demand response program planning. The challenge in designing a RASS is to construct a sample selection and data collection methodology that maximizes precision of resulting estimates, minimizes bias and minimizes respondent burden all while adhering to project scheduling and funding constraints.

There were numerous challenges that the DNV GL/Seminole team faced in designing and conducting the 2016 RASS – primarily due to the limited project budget. The following paper will present the methodologies used to design the study, collect the data and create final estimates and their precision. Emphasis will be placed on summarizing project challenges and their respective solutions. These strategies and lessons-learned are valuable for researchers interested in any type of low-cost study of their residential customers.

Introduction

A comprehensive examination of energy usage in a population should address two fundamental concepts: **quantity** i.e., how much energy is being used by the population, and **source** i.e., what is the energy being used for. There are exceptions but in general, measuring the amount of energy consumption (quantity) in a population, at least in the United States, is not too difficult. Utilities, for example, meter their customers and will have billing data that contains energy consumption information for each of their accounts. The much more difficult and highly variable concept to address is what the energy is being used for (source). Addressing and understanding this latter concept is necessary to identify feasible and creative ways of reducing energy consumption in a population.

One method that is used by numerous utilities, cooperatives and other interested parties to identify sources of consumption in a *residential* population is to conduct what is referred to as a **residential appliance saturation survey (RASS)**. A RASS contains various questions that gather data on topics including: occupancy status (renter/owner), structure of the house (e.g. vintage, square footage, number of floors), demographics of the household occupants (age group, income), heating/cooling systems (heat pump, furnace, age/size of systems), lighting (lamp type, hours of use) and appliances (refrigerators, washers, dryers, electronics).

A RASS is generally administered to a sample of the population, in contrast to the entire population, to minimize data collection costs and burden. RASSs are administered using numerous combinations of data collection modes including mail, email, telephone, web and in-person visits.

In 2016, DNV GL conducted a RASS on behalf of Seminole Electric Cooperative Incorporated (Seminole) for each of its nine Distributor Member Cooperatives' service territories.[1] Seminole is a wholesale generation and transmission provider to the Member Cooperatives. One of the services Seminole provides is to conduct an appliance saturation survey (i.e. a RASS) at least every five years. The survey was conducted in-house since 1980 and Seminole sought to outsource this project for the first time in 2016. Seminole ultimately teamed with DNV GL based on competitiveness in cost, expertise and client referrals.

The historic series of Seminole appliance saturation studies are somewhat unique, compared to RASSs implemented by other organizations, in that they enabled researchers to not only measure the varying sources of consumption among each of the nine residential populations but they also enabled researchers to measure change in consumption as well as differences in consumption sources between the cooperatives.

Change in Consumption
This was achieved in 2016 by comparing results from one study with results from previous Seminole RASSs conducted among the same population.

Differences in Consumption Sources Between Cooperatives
Comparability of estimates between cooperatives was achieved in 2016 by using the same data collection instrument, same data collection, sampling and weighting methodologies.

The following paper summarizes the methodology and results from the 2016 data collection effort. Discussions will touch on the sample design, development of the instrument, identification and implementation of the data collection methodology within the prescribed project cost and schedule constraints, methods used to expand the survey results back to the population (sample weighting), and results from the study. The methodologies used to address these study features will be of interest to many readers, however the primary focus in the discussion below will be on summarizing the challenges we faced in conducting the study and our solution for addressing these challenges.

Sample Design

The success of any study initially depends on the design and sample selection methodology used to get the study started. The importance of a well-crafted sample design cannot be underestimated. The ability to generate precise, unbiased estimates from collected data is highly dependent on the methodology used to select the initial sample. While additional factors such as an effective data collection process and a well-understood instrument also affect precision and bias, they are less important without a solid start to the study offered by an efficient, optimized design.

A scientifically defensible sample design of any population needs to address various features, and for the most part, these features were easy to identify with the 2016 Seminole RASS. For example, the **target population** for this study was all residential customers in 2016 within each of the Member Cooperatives' service territories. The **primary subgroups of interest**–these are the groups that estimates are desired "by"–were the nine Member Cooperatives. Nonresponse and the potential for **nonresponse bias** was a concern during the design, this is something we sought to minimize (both unit and item nonresponse.) The **sample frame**, which needs to identify as cleanly as possible each member of the target population, was developed from billing data records obtained from each of the Member Cooperatives. It

[1]For confidentiality purposes, the specific names of the nine Members have been omitted from this paper. Researchers interested in 2016 RASS results for a specific Member should contact Seminole Electric Cooperative.

was known at the onset that the project **budget**, and to a lesser extent the **schedule**, were extremely aggressive on this project.

After careful and thoughtful discussion, and given the desire for generating precise estimates with a limited budget, the team decided the primary mode of data collection would be via web using email messages as the primary method of eliciting a response from selected households. This meant the 2016 RASS sample would be selected from among those households in the sample frame with an email address. We could have contracted with a vendor to obtain email addresses for those households on the sample frame that didn't have one, but for cost and scheduling reasons this was not pursued.

Our decision to concentrate the data collection effort among those households with an email address on the sample frame was not made lightly. There are strong pros and cons affiliated with this decision as noted in Table 1.

Table 1. Pros and cons associated with decision to only sample households with an email address

Pros	Cons
• Limiting the data collection effort to just those with an email address and eliciting a response to the selected households only using email meant the variable portion of the data collection costs would be minimal. In other words, we could get "more" responses for the same budget. • The larger respondent sample size meant the precision of estimates would be greater.	• We suspected nonresponse bias could be significant since people would likely view an unrecognizable email from their cooperative as simply "spam." • There is a potential for target population coverage bias. This potential exists when parts of the target population are, accidentally or intentionally, omitted from the sample selection process. In this case, households with no email address would be omitted from the sample selection process.

The biggest concern was the coverage bias. Steps were taken to minimize coverage bias such as:

- We gave non-selected households that learned about the RASS through some external correspondence two methods to respond to the study. These "volunteers" could respond to the study either by filling out the web questionnaire as a "visitor" or by requesting a paper version of the questionnaire.
- DNV GL prepared survey marketing materials for the Member Cooperatives to use in their e-newsletters, utility magazine advertisements, and social media pages. This maximized the chance of volunteers learning about the study through an external source.
- The respondent sample was statistically weighted to the target population within each Member Cooperative using kWh consumption and a wide variety of neighborhood characteristics that are likely correlated with survey items of interest. Controlling for these attributes in the sample expansion process would therefore decrease the coverage bias of resulting estimates from the study.

Ultimately, however, the potential coverage bias implications of the decision to use web as the only primary mode of data collection are uncertain.

The sample selection process itself proved to be rather minimal and simplistic, primarily because of the small incremental costs associated with eliciting a response from a larger number of households on the sample frame with an email address. For smaller Member Cooperatives, all households with an email address on the frame were selected for the RASS. For the larger Members, a stratified simple random sample of households was selected within strata defined by normalized annual consumption. The goal was to select roughly 10,000 households within each Member Cooperative.

Sample Frame Development

One of the biggest and costliest challenges associated with this study was the development of the sample frame. As noted previously, the sample frame was constructed from billing data obtained from each of the nine Member Cooperatives. Table 2 summarizes the total number of households on the frame, the initial sample counts and the percent of the target population with an email address. The percent of the target population with an email address varied considerably between 11% and 59%. This suggests the potential for coverage bias may also vary considerably between each service territory.

Table 2. Summary of frame and sample

Group	Member A	Member B	Member C	Member D	Member E	Member F	Member G	Member H	Member I
Total Households	150,144	29,481	11,178	31,958	174,519	24,488	11,864	48,740	181,182
Percent with Email	11%	28%	33%	37%	33%	42%	32%	59%	46%
Selected Sample	9,483	6,262	2,938	11,181	9,794	7,019	3,535	9,227	9,961

Below is a summary of challenges and lessons learned while developing the sample frame. Many of these won't come as a surprise to analysts familiar with billing data originating from various utilities:

- Address information for each household was requested; however, in some cases the mailing address was provided instead of the "physical location" of the household. Physical location was needed to merge neighborhood characteristics onto the sample frame (this is discussed below.)
- "Address" information was requested, and in one instance the only data item supplied was street address. City, state and zip code were missing.
- Consumption information was requested for the previous 12 months. However, meter read-dates were not always provided. This made creating a normalized annual consumption (NAC) for each household difficult because the exact time-period consumption data corresponded to was not always known.

The lessons learned from this experience are somewhat obvious: the analysts on our team learned their written request for billing data to the Member Cooperatives needed to be as clear as possible and explicitly state exactly what is needed. Additionally, providing some context on why certain data items are needed would have also likely helped in their file preparations. For example, it was probably not clear that street address, city, state and zip code were critical data items in an email survey as they were needed by our analysts to merge neighborhood information onto each record. Member Cooperatives also probably didn't realize our desire to create a NAC for each household and that meter read dates were necessary to do this.

Earlier we noted physical location address was used to merge neighborhood information onto each household record on the sample frame. To achieve this, the study team first geocoded each household on the frame to obtain the latitude and longitude associated with the physical location address. These geographic identifiers were then used to identify the U.S. Census block group, tract and zip code tabulation area for each household. These variables were used to merge on neighborhood information from the 2010-2014 American Community Survey (ACS). The smallest geographic areas for which neighborhood-level estimates are available are block groups. Because of low precision, some ACS estimates are only available at the tract and zip code tabulation area. A discussion of the neighborhood variables of interest in this RASS is included in the sample weighting section below.

Table 3 provides a summary of the percent of frame accounts that were successfully geocoded. For example, 92.5% of the Member Cooperative A households were geocoded. This means latitude/longitude was identified for 92.5% of their residences and therefore tract and block-group-level ACS neighborhood information was merged onto the frame for this group. For the remaining 7.5%, zip code centroids or neighborhood averages computed at the Member Cooperative-level were used. The most informative neighborhood information would be that obtained at the block group and tract level, so the estimates in Table 3 suggest “good” neighborhood information was identified for most of the frame households.

Table 3. Percent of residential records on sample frame that were geocoded

Member A	Member B	Member C	Member D	Member E	Member F	Member G	Member H	Member I
92.5%	75.3%	75.5%	74.8%	68.5%	72.2%	89.2%	90.9%	56.9%

The percentage of households successfully geocoded varied considerably between Members Cooperatives, ranging from 56.9% to 92.5%. This reflects the varying quality of the address information received from each Member Cooperative. PROC GEOCODE from SAS® (2016) was used to geocode each address. Their procedure proved very efficient for this application. Had the budget and schedule permitted, we could have used additional sources to geocode the residual non-geocoded set from PROC GEOCODE.

Instrument Development

The 2016 RASS data collection instrument was based on Seminole’s 2011 Residential Consumer Survey. This was intentional, to allow for consistent trending analysis. Minor modifications were deliberate and included:

- Formatting questions to accommodate a web- versus paper-based survey. These changes were needed because the 2011 study was administered via a paper version of the questionnaire and the primary mode of data collection for the 2016 study was web.
- Adding a new battery of questions to capture data on solar installations. In general, any additions to the survey were carefully considered to avoid unnecessary respondent burden, as well as potential survey fatigue that can frequently lead to biased responses.
- Updating response options to reflect technologies that are more commonplace in the current marketplace such as ductless mini/multi-splits.
- Addition of measures and granularity of response options in anticipation of potential downstream uses of the data such as an energy efficiency potential study.

An iterative, collaborative, and inclusive stakeholder review process ensured that the final data collection instrument addressed Seminole’s and the Member Cooperatives’ collective and individual data needs.

Once the survey content was finalized, the DNV GL programming team translated the paper questionnaire into a web survey using a service provided by WorldAPP (2016). The web team encountered some technical challenges translating the paper document into a web data collection tool. For example, questions needed to be reformatted, modified, or even broken into multiple parts to accommodate system limitations or requirements. To some extent, these challenges were expected.

After programming was completed, systematic testing and quality assurance checks were conducted to verify the web based tool was collecting and storing data as intended. This process took longer than anticipated. It is easy to underestimate the amount of time testing and quality control can consume, and it is not a step in the process that should be rushed. In addition to performing content review for each individual response option, the skip pattern and question type (single versus multiple response) were verified. This process was tedious yet critical to the success of the project.

Data Collection

Based on stakeholder budget and time constraints it was decided to conduct data collection primarily via an online web survey. A limited number of Member Cooperatives offered paper surveys at their customer service centers and as noted earlier, some respondents "volunteered" to take the survey and responded to the web instrument as a visitor. However, these volunteers ultimately represented less than 0.2% of the total responses. On-line data collection for a RASS has several desirable characteristics that include:

- Online surveys produce **faster results at a lower cost** than telephone or mail based surveys.
- The **quality of the responses improves** because respondents can control the time and pace of their responses. And the instrument can include visual cues to aid in understanding energy efficiency technologies such as illustrating the differences between CFL and LED bulbs.
- Data quality from well-programmed web surveys is often better because the online instrument can include **automatic skip patterns and unambiguous data entry requirements**. For example, the instrument can be programmed to accept only reasonable numeric values to questions that require it (e.g. zip code).
- Skip patterns improve data quality and shorten the time required for completing the survey, therefore **minimizing the burden** on respondents.
- An online instrument **minimizes item nonresponse and unwanted multiple responses** by preventing respondents from skipping questions and assuring respondents only enter a valid single response for single answer questions.

Initially a two-pronged survey distribution approach was discussed: (1) sending an email to a sample of households with a hyperlink to the survey and (2) marketing the survey via a bill insert that included the survey web address to further encourage "visitor" responses. During the planning process, each Member Cooperative investigated the lead time necessary to coordinate and schedule a bill insert. Without exception, it was decided that a bill insert would not be possible considering the aggressive data collection schedule for this RASS. It was agreed that the survey would be distributed via email only. A survey invite e-mail template was crafted and distributed to the Member Cooperatives who were provided the option to customize it.

Data was collected over a four-week period and most responses were received within five days of the initial email distribution. Throughout the data collection process the evaluation team delivered system generated reports that provided Member Cooperatives with their respective survey response rates. These reports helped each Member decide on whether to email reminders to nonrespondents to further increase their respondent sample (and reduce nonresponse bias). Due to healthy response rates observed from the initial email contact, and to avoid over-burdening their members, no Member Cooperatives elected to send a reminder.

Table 4 shows the final number of respondents and the final response rates achieved. Response rates varied from 13.2% to 26.6% across the Members. Considering the data collection methodology consisted primarily of one email with no reminders, we felt the higher than anticipated response rate was a success for this study.

Table 4. Respondents and response rates

Group	Member A	Member B	Member C	Member D	Member E	Member F	Member G	Member H	Member I	Total
Respondents	2,023	1,353	653	2,742	2,164	1,497	1,219	649	2,651	14,951
Response Rate	21.3%	21.6%	22.2%	24.5%	22.1%	21.3%	13.2%	18.4%	26.6%	21.5%

Expanding Study Results to the Population – Sample Weighting

The analysis of the Seminole RASS data began by aggregating all the final output files from the various web instruments into one main analysis file. Some basic checks were performed, such as identifying and removing duplicate records. In this study, we required that a respondent answer the questions on basic household characteristics as well as the lead-in question about cooling systems in their home for a response to be considered a final "completed respondent".

After the set of respondents was finalized, the next step in the analysis process, which is often one of most difficult, was to develop an appropriate expansion factor for each household that responded to the survey so that results from the study can be expanded back to the original population of interest. This numeric expansion factor is often called a **sample weight**.

In a typical probability survey where a sample is drawn from a frame that closely matches the target population of interest, the sample weight begins with the inverse of the probability of selection. One or more adjustment factors are then applied to the weight to account for survey nonresponse and slight variations between the frame and the target population of interest. These are referred to as a **nonresponse weight adjustment** and a **post-stratification weight adjustment,** respectively. In this study, for a few of the larger Member Cooperatives there was some sampling done. However, ignoring nonresponse for a moment, using the inverse of the probabilities of selection to create a sample weight would not be sufficient for this study, primarily because the study was implemented among those that had an email address on the billing data sample frame and those with an email address represented only a small portion of the target population (see Table 2). The target population was all residential households for each of the nine Member Cooperatives in their entirety, not just those with an email address.

There were numerous challenges in creating a sample weight in this study, and most of them are similar to what would be encountered in any study where a sample is selected from a small subset of the population such as those with an email address. These challenges include:

- **Undercoverage Bias.** The sample selection and data collection methodologies were designed to collect data from a purposely identified subset of the entire target population. In our study, this was the population that had an email address on the billing data sample frame. However, to the extent possible, we wanted the resulting sample to expand back to the entire household populations and not just those who had an email address.
- **Sample Selection Bias**. Our team acknowledged at the onset that estimates may be biased due to the primary mode of data collection used to gather responses (web) and the way the sample was selected (only those with an email address were selected). But to the extent possible, we wanted the results of the RASS to expand back to the entire household population in a manner that minimizes this bias.
- **Limited Information Known About the Target Population.** Another key challenge was the lack of information known about the entire household target population for each of the Member Cooperatives. The sample frame was built from billing records obtained from each Member. As with most utilities, billing data provides two solid pieces of information on each household: previous consumption patterns and address/location information, and that's about it.
- **Self-Selected Respondents.** It was noted earlier that one feature we included in the data collection process was a method by which non-solicited households could respond to the RASS by going to the web instrument and filling out the questionnaire as essentially a "visitor". This feature of the data collection process was included to at least partially account for the sample selection bias noted earlier because it's a way of allowing households that did not have an email address a means of responding to the study. Very few responses were obtained from this set. Nevertheless, respondent burden should never be "wasted" in a study and it was therefore our desire to include

these respondents in the final analytic file along with an appropriate sample weight. Zip code was collected for these respondents. Average consumption data for the zip code was assigned to these households so that they at least had the same frame information that the selected sample did.

To address these challenges while maintaining the scientific integrity of the sample selection process as much as possible, we created a sample weight for each completed household that consisted of two factors:

1. **Inverse of the Probability of Selection**. The probability of selection was set equal to 1.00 for the self-selected respondents and those selected from Member Cooperatives where all households with an email address were solicited for the data collection.
2. **Target Population Calibration Weight Adjustment**. This adjustment was designed to address as many of the previously noted challenges as possible.

The much more important adjustment factor of the two was the calibration weight adjustment. This adjustment was developed by fitting a constrained, generalized exponential model. In our application, the dependent variable in the model is the response indicator and all households on the sample frame that were not considered final RASS respondents were considered "nonrespondents" in the modeling/calibration process. Parameters in this model were estimated using calibration equations that were constructed so that the sum of the model-predicted weight adjustment among respondents equals the entire population across the independent variables used in the modeling process. The constrained, generalized exponential model and calibration estimation process are discussed in Folsom and Singh (2000) and Witt (2009).

Research Question: so, what does the calibration weight adjustment methodology mean for this RASS and how does this address the challenges noted earlier? To answer this, consider the weight adjustment methodology first. Suppose, as an example, there are two pieces of data available for every household (HH) on the sample frame: normalized annual consumption and the percent of housing units in the neighborhood that are owner occupied (from the ACS data). Suppose we create six yes/no indicator variables from these two pieces of data defined as follows:

Normalized Annual Consumption (NAC)
1. HH in upper quartile (Yes/No)
2. HH in middle two quartiles (Yes/No)
3. HH in lower quartile (Yes/No)

Percent of Owner Occupied Housing Units
4. HH in a neighborhood with < 50% (Yes/No)
5. HH in a neighborhood with 51%-75% (Yes/No)
6. HH in a neighborhood with > 75% (Yes/No)

Suppose a calibrated weight adjustment is created using the generalized exponential model with the above six variables as explanatory variables. Then the model will yield an adjusted weight: the weight will differ between respondents and most importantly, each respondent will only get one adjusted weight. And the sum of this weight for households in each of the above six groups will exactly equal the correct population total from the sample frame. Therefore, the generalized exponential model yields a sample weight for each respondent that satisfies the property that the sum of the weights across respondents is perfectly calibrated to the target population totals across all six categorical variables included in the model.

Returning to the research question above, this weight adjustment approach was used to lessen the coverage bias as much as possible by creating an expansion factor using as much of the frame information as possible. Specifically, separate exponential weight models were estimated for each of the nine Member Cooperatives that included independent variables such as categorized NAC and 73 categorized versions of 12 American Community Survey neighborhood variables that included median number of rooms in housing units (HUS), percent owner occupied and distribution of HUs by year structure built.

Table 5 summarizes a few results after the calibration weight adjustment was created and applied. This table shows estimates over the entire Seminole target population, i.e. results summed across all Member Cooperatives. For example, this table shows that 50.2% of all households in the Seminole territory reside in neighborhoods where 6%-15% of the families are living below the poverty threshold. Looking at the completed respondents with no weight adjustment, this percentage is 45.0%, so the sample is under-representing this group slightly. And after applying the weight adjustment the weighted estimate from the sample is 50.2% - exactly equal to the target population percent.

Table 5. Comparing unweighted and weighted distribution of eligible population over all territories

Group	Target Population	Unweighted Sample	Difference	Weighted Sample	Difference
Weather Normalized Annual kWh					
Low (0 - 8,034)	18.9%	18.9%	0.0%	18.9%	0.0%
Medium-Low (8,035 - 11,999)	20.1%	20.3%	-0.2%	20.1%	0.0%
Medium (12,000 - 15,358)	20.2%	20.7%	-0.5%	20.2%	0.0%
Medium-High (15,359 - 19,566)	20.5%	19.5%	1.0%	20.5%	0.0%
High (19,567 Plus)	20.3%	20.6%	-0.3%	20.3%	0.0%
TENURE, Universe=Occupied housing units: Percent owner occupied					
Low (0%-74%)	24.7%	23.5%	1.2%	24.7%	0.0%
Medium (75%-89%)	49.9%	49.7%	0.2%	49.9%	0.0%
High (90%-100%)	25.4%	26.8%	-1.4%	25.4%	0.0%
POVERTY, UNIVERSE=Families: Below Poverty					
Low (0%-5%)	25.7%	28.6%	-2.9%	25.7%	0.0%
Medium (6%-15%)	50.2%	45.0%	5.2%	50.2%	0.0%
High (16%-100%)	24.1%	26.4%	-2.3%	24.1%	0.0%

In conclusion, can we confidently state that the weight adjustment addressed the potential biases in the estimates that were noted earlier in this discussion, such as undercoverage bias or sample selection bias? The answer of course is absolutely not - we cannot be certain this removes the bias. Some additional data collection efforts that target the missed population would need to be done to confirm this. But to the extent the bias is related and correlated to the frame information we have, such as annualized consumption and neighborhood characteristics, we confidently believe we reduced the bias as much as we could given the constraints of the design, budget and schedule of this study.

Comparing Estimates with Estimates from Prior Years

One key point of interest to Seminole and its Member Cooperatives was comparing the 2016 RASS results to data collected from past surveys conducted in-house since 1980. DNV GL added enhancements to Seminole's previous survey efforts that included:

- The 2016 study resulted in a much larger sample of participants relative to Seminole's past surveys.
- The 2016 results were weighted using a calibration approach designed to reduce coverage bias.
- A multiple mode of data collection was applied (primarily web but paper was offered). Previous studies were conducted using a paper instrument only.

The consensus among the Members was that these enhancements likely improved the quality and reliability of the estimates generated from the 2016 study. But the nagging question still existed - how do these estimates compare to previous studies of the same populations. In other words, how do these enhancements effect the trend and historical comparability of estimates. Astonishingly, most results aligned almost seamlessly with the evolution of appliance saturation and equipment efficiency depicted in Seminole's past surveys. Exhibit 1 shows trends in Seminole service territory types of homes and primary air and water heating systems as an example. One can see the 2016 results seem to follow the same pattern suggested by survey results since the 1980 Seminole study.

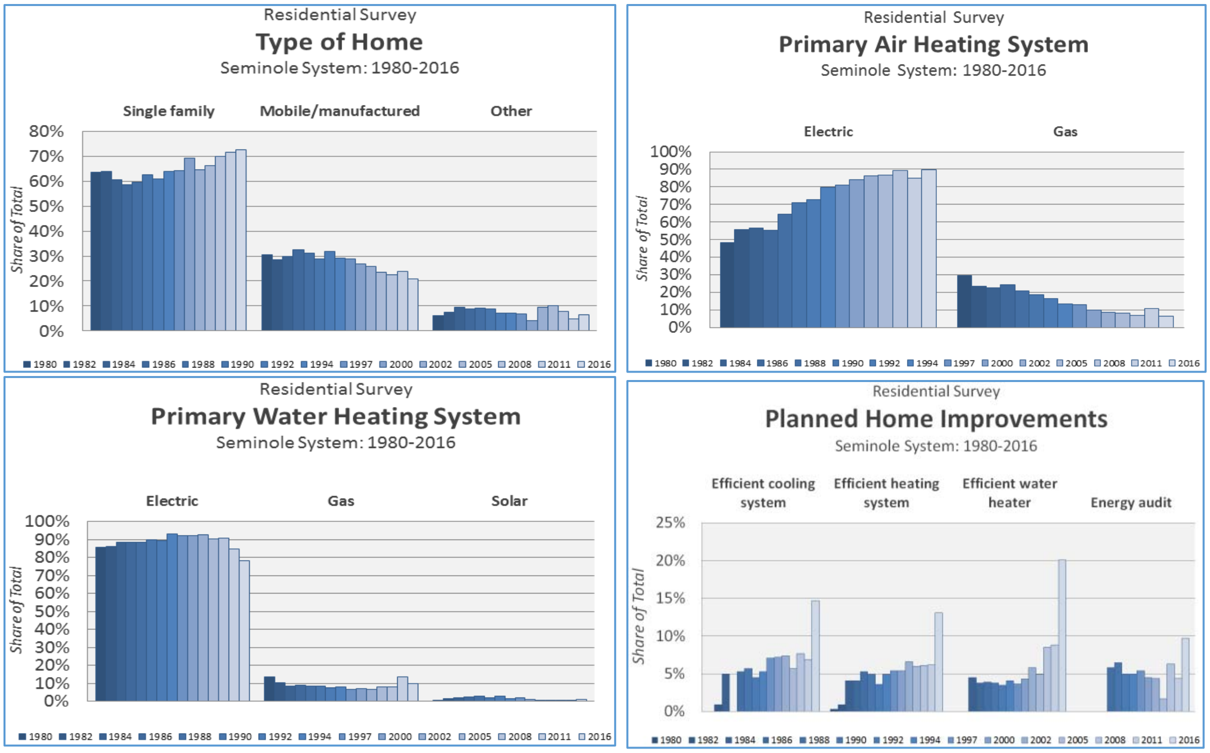

Exhibit 1. Trend estimates between 1980 and 2016

There were instances where the 2016 survey results indicated a dramatic difference compared to previous studies. For example, note the last graph in Exhibit 1 that shows the percentage of households with a planned home improvement. The 2016 estimates are more than double what was observed in the 1980-2011 studies. One of the unknowns in this trend analysis is the affect the 2016 data collection and weighting enhancements have on trends such as this. In other words, is the large increase in the percent of households with a planned home improvement in 2016 due to differences in the 2016 methodology or is this a real change in the population? This is something the 2016 results are not be able to answer.

Comparing Estimates Between Member Cooperatives

As noted earlier, one of the major features of the Seminole RASS was that the same type of data, using the same instrument, were collected from a sample of residential customers in each of the nine Member Cooperatives. This equivalency in the type of data collected was done intentionally. One of the major purposes of this study was to compare estimates across service territories.

One challenge we faced related to simple data presentation. We wanted to show researchers tables of estimates for each of the nine Member Cooperatives, and from a statistical perspective, it was important to convey to readers that each estimate has some sampling variance associated with it. For example, we found the estimates of the percent of households with one or more CFL bulbs ranged from 75% to 80% across service territories. Looking at the extremes, just because one Member Cooperative has an estimate that is five percentage points larger than another does not mean the difference is noteworthy from a statistical point-of-view. It may be the observed difference is due to the sampling error associated with each estimate. The study team's challenge was how to flag differences that were statistically significant among nine different estimates in a manner that was easily digestible to an audience that may not be very familiar with the concept of sampling variability and sampling error.

A table was created for each survey question that showed the estimates by Member Cooperative, the respondent and nonrespondent sample size and the margin of error associated with the estimates. We noted in each table that the "margin of error is the 90% confidence interval half width. The 90% confidence interval associated with an estimate is the estimate plus/minus its half width." Behind the scenes, the margin of error was computed using a t-statistic along with an estimate of the standard error of the estimate. The standard error of an estimate is the square root of its variance estimate. And variance estimates were computed using the **Taylor Series Linearization** method. In general, the Taylor Series Linearization process for estimating variances accounts for the complex design features that are often found in survey samples, such as stratification, clustering and/or unequal weighting. So, this variance estimation process is appropriate for estimates from this study.

To provide the reader with some information on whether a difference between estimates was statistically significant, we used superscripted letters besides each estimate to indicate whether an estimate was statistically significantly different from another estimate in the same row. This is illustrated in Table 6. For example, the table shows the estimate for Member A (34.5%) is significantly different from the estimates associated with Members C through I but not Member B. This also shows the patterns of significance are not always obvious. For example, the estimates for Member C is significantly different than the estimates for Member G and I but not H, even though the estimates for H is between G and I. This phenomenon can be due to a variety of factors. The respondent sample sizes are not the same between Member Cooperatives. We're looking at table that displays statistical significance associated with a large number tests and therefore some are expected to show incorrect inferences. Additionally, the Taylor Series variance estimate used in these tests is subject to variance itself, and this could affect the results of any statistical test. Regardless of the shortcomings though, this is an effective way of conveying to the reader the likelihood that one estimate is different from another given the lack of precision due to sampling variability associated with the estimates.

Table 6. Percent that have done/plan to do weather improvements to their home in the next 2 years

Member A	Member B	Member C	Member D	Member E	Member F	Member G	Member H	Member I
34.5%^{cdefghi}	37.0%^{efghi}	37.7%^{agi}	38.9%^{a}	40.6%^{ab}	41.0%^{ab}	42.0%^{abc}	43.2%^{ab}	45.0%^{abc}

Note: A superscripted a, b, c, etc. indicates the estimate is significantly different from the estimate in column a, b, c, etc. at the .10 level of significance.

Looking at estimates between independent groups often leads analysts to wonder whether any differences observed can be explained by differences in the populations. This is often thought of when one is comparing estimates between two populations as well as when one is comparing estimates for the same population at two different points in time. For example, considering the estimates in Table 6, a larger percent of customers in Member I may be more interested in weatherization, compared to Member A, because there is a larger percentage of older homes in the Member I service territory and not necessarily because there is more interest in reducing energy for "admirable" reasons. We may have also found, for example, that the 45.0% estimate for residents in Member I to be a statistically larger estimate than what was found for Member I several years ago, but this could be because the distribution of the population by working status has changed between the time periods and homeowners now have additional funds available to invest in weatherization improvements.

There are various methods one can employ to understand the "why" part of any difference observed in a study. Some methods are more direct, for example a survey instrument could include additional questions that ask about reasons for doing an activity or having a particular attribute.

Differences in these reasons between groups might shed some light on why the prevalence of the activity is different between groups. The one problem with this method is that analysts generally don't know which differences are worth investigating until the differences are seen after the study is complete.

There are also various statistical methods one can use to explain differences. For example, one could model a dependent variable against a set of independent variables and look for differences in the significance of one or more independent variables between groups.

Another method statisticians use to compare estimates is to compute **standardized estimates** (see for example, Witt and Spagnola, 2009). Standardization is a technique used to account for differences in population composition which may have an impact on estimates of an outcome measure (Kalton, 1968; Konijn, 1973). When dealing with survey data, standardization can be thought of conceptually as creating an adjustment to the final sample weights so that the distribution of the reweighted sample in groups of interest equals some fixed distribution. This fixed distribution is often referred to as the **standardization population**. The standardization population can be obtained from an outside source, or is often estimated from the entire sample without regard to the group(s) being considered. The standardized estimate (sometimes referred to as the adjusted mean) can be interpreted as the estimate that would have been obtained if the group exhibited the distribution of the standardizing population with respect to those characteristics being controlled for, all other things being equal (Little, 1982).

To illustrate the standardization methodology, consider four standardization variables and define the standardized population as the entire sample. The four variables are structure type (single family home, mobile home, apartment/condominium), year-round residence indicator (yes or no), age of home (0-15 and 16+ years-old) and size of home (< 1,200; 1,200-1,999 and 2,000+ square feet).

Table 7 shows the percent distribution of customers in the nine Member Cooperatives by the year-round residence indicator, as an example. Notice the percent of customers that live at their residence year-round varies from 73.5% to 96.4%, and the estimate for the entire pooled population is 89.4%. With the standardization methodology, effectively a weight adjustment is created for each respondent so that the new weighted distribution within each service territory would equal the total estimate. For example, currently 73.5% of the residents in Member G are year-round residents. This would change to 89.4% with the standardization weights.

Table 7. Percent of households that are year-round residents

Member A	Member B	Member C	Member D	Member E	Member F	Member G	Member H	Member I	Total
92.5%	84.0%	81.8%	84.3%	83.2%	90.2%	73.5%	96.4%	94.0%	89.4%

Table 8 illustrates the difference between the final nonstandardized estimates and standardized estimates of the percent of households with one or more interior LED bulbs, when the nine populations are standardized to the four variables noted above. Notice the "current" final estimates range from 40.4% to 54.8% but the range of the standardized estimates is less at 43.4% to 52.0% suggesting some of the differences observed in the final estimates were likely attributed to differences in the Member Cooperatives' populations by the four standardization variables. Also, note that many of the significant differences seen in the final estimates are no longer significant with the standardized estimates. For example, the final estimate for Member A is different from Members E and F, but this difference is not significant in the standardized estimates.

Table 8. Percent of households with one or more interior LED bulbs

Estimate	Member A	Member B	Member C	Member D	Member E	Member F	Member G	Member H	Member I
Nonstandardized	40.4%^{efghi}	41.0%^{fghi}	43.1%^{fghi}	43.2%^{ghi}	44.9%^{aghi}	47.5%^{abcghi}	51.5%^{abcdefi}	51.6%^{abcdefi}	54.8%^{abcdefgh}
Standardized	43.8%^{ghi}	43.4%^{ghi}	44.9%^{ghi}	44.6%^{ghi}	45.1%^{ghi}	46.9%^{ghi}	52.0%^{abcdef}	50.4%^{abcdef}	51.5%^{abcdef}

Note: A superscripted a, b, c, etc. indicates the estimate is significantly different from the estimate in column a, b, c, etc. at the .10 level of significance.

References

Folsom, R.E. and Singh A.C. (2000) "The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and post-stratification." *Proceeding of the* 2000 *American Statistical Association, Survey Research Methods Section*, pp.598-603.

Kalton, G. (1968). Standardization: A technique to control for extraneous variables. *Applied Statistics, 23*, 118-136.

Konijn, H.S. (1973). *Statistical theory of sample survey design and analysis*. London: North-Holland.

Little, R. J. A. (1982). Direct standardization as a tool for teaching linear models for unbalanced data. *American Statistician, 36*(1), 38-43.

Witt, M., & Spagnola, K. (2009) Using predictive marginals to produce standardized estimates. Proceedings of the American Statistical Association, Section on Survey Research Methods, 4718-7432.

Witt, M. (2009). Overview of software that will produce sample weight adjustments. Proceedings of the American Statistical Association, Section on Survey Research Methods, 3009-3023.

SAS (2016). See https://www.sas.com/en_us/home.html. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

WorldAPP (2016). See https://www.worldapp.com/ .