1 Background

Throughout November 2020, the Board collected data on water system financial impacts and household water bill debt accumulation during the COVD-19 pandemic from a sample of public water systems. The survey responses will allow the Board to formulate statewide estimates for water systems that may be facing financial crisis and the number of households with water bill debt, including the level and geographic distribution of debt. The Board is collecting this data to inform policymakers of options for financial assistance and emergency response for water systems and households experiencing economic hardship.

Different surveys (questions and survey design) were completed, one for community water systems (CWS) serving up to 10,000 connections and another for CWS serving more than 10,000 connections, referred to in this report as “small” and “large” systems, respectively.

2 Survey Design and Analysis

2.1 System Sample Stratification

Water systems serving less than 10,000 service connections were binned according to service connection using the Jenk’s Natural Breaks method to reduce variance. The bins were determined to be as follows:

- Bin A: Less than 1,009 service connections
- Bin B: 1,009 to 3,389 connections
- Bin C: 3,090 to 5,867 connections
- Bin D: 5,868 to 10,000 connections

Within these four discrete strata, water systems were randomly sampled. The sampling design strata are shown in Figure 1.

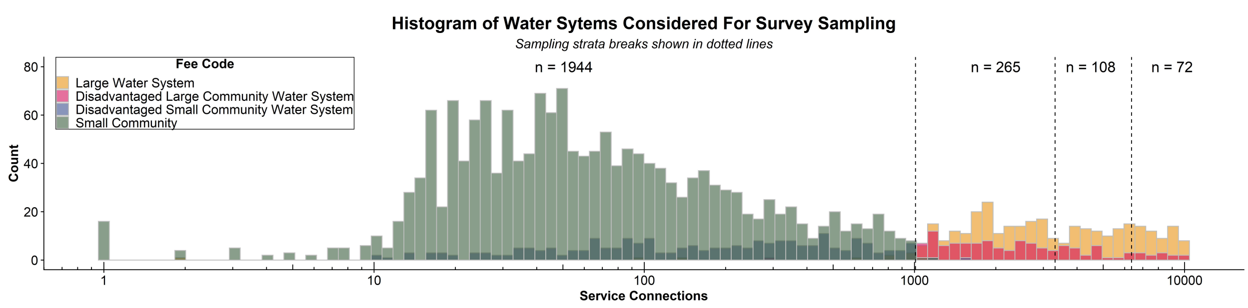

Figure 1: Distribution of <10,000 service connection water systems by service connections, annotated with natural breaks

Systems with greater than 10,000 service connections were sampled separately.

2.2 Survey Questions

Both surveys asked for the following from all systems in the sample:

- Total debt from all delinquent accounts and number of delinquent accounts
- A per-zip code and per-debt level breakdown of delinquent accounts

The small system survey questions included:

- 2019 and 2020 system revenue and expenses
- Cash reserves: restricted, unrestricted, and total
- Anticipated number of months before financial assistance is required

The large system survey questions included:

- Debt solely related to drinking water
- Repayment plan options
- Late fee charges, if applicable

3 Data Collection and Survey Completion Rate

3.1 Survey distribution

In order to maintain a level of data consistency with regards to the response type (e.g. ensuring that numerical answers are provided in numerical format), a survey template was created for each of the two survey types in the form of an Excel document. Using pre-existing information on the system public water system ID (PWSID), system name, and billing cycle as provided in the 2018 Electronic Annual Report, staff generated a unique survey form for each system in the sample list with the Python *openpyxl* library.

The survey takers were instructed to only provide information in the designated cells within the survey form; these cells were formatted with data validation to ensure that the response type was in the appropriate format. While staff attempted to obtain results that were as complete as possible, there were a number of limitations (software, COVID impacts to staff, insufficient time to gather necessary data) that prevented complete or sufficiently detailed responses. Under those circumstances, limitations were noted in the provided comment boxes.

3.2 Small water systems

The total set of community water systems with less than 10,000 service connections (2,661 in total) were binned (see “System Sample Stratification”) and 510 water systems were selected for the sample accordingly. Each Division of Drinking Water (DDW) field office directed staff to call the systems on the sample list that fell within their district, along with any systems that fell under the jurisdiction of local LPAs. In total, 406 surveys (428 systems) were completed, with 7 surveys representing multiple systems.

3.3 Large water systems

150 large water systems out of the total 223 large community water systems were initially chosen to be part of the survey effort, with the selection process being based on system size and region. On November 9, 2020, Water Board staff sent each system an e-mail with the survey attached, as well as an FAQ sheet and instructions. Over the next week, staff called each system to confirm receipt of the survey and to invite them to a Q&A workshop aimed at assisting systems complete the survey. The workshop took place on November 18, 2020 and 163 people signed up to attend the workshop (some systems sent multiple representatives).

The large system survey was initially due on November 25, 2020, although December 4th extensions were granted in order to guarantee a sufficient survey response rate. During the extension period, Water Board Staff continued outreach efforts via phone and email to all systems who had not yet submitted the survey. In total, 131 surveys (151 systems) were submitted, 5 of which were volunteer surveys (not in the original sample of 150). 3 of the returned surveys were submitted on the behalf of multiple systems. 14 systems of the original 150 did not submit a survey – 2 systems replied that they were unwilling to participate in a voluntary survey, and 2 systems replied that they would submit a survey but did not follow up. Finally, 1 system suffered a COVID-19 outbreak on staff that prevented them from completing the survey. Only 2 systems never responded to any outreach via phone or email.

3.4 Survey Completeness

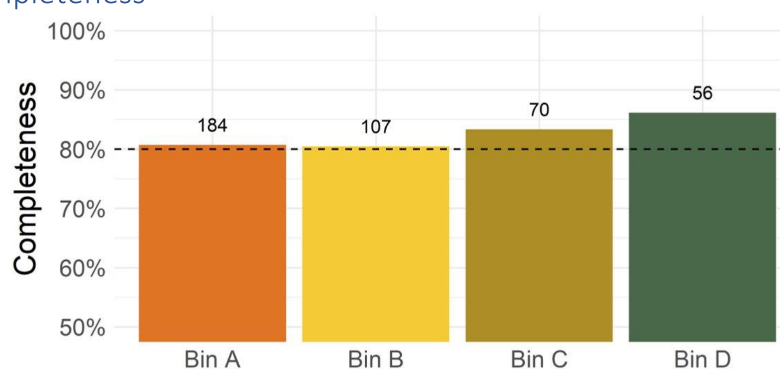

Figure 2: Small survey completion rate by bin

All four of the sampling bins for the small systems, as well as the large system sample, had a response rate of over 80%, although not all questions were answered by respondents.

4 Survey Analysis

Once data collection was finalized in December, the results from individual surveys were extracted programmatically using Python; only results that were provided in the designated cells were extracted. The results were compiled into a final Excel document with separate tabs for the large and small survey questions, although the zip code table was combined into a single output table with an additional column denoting whether the corresponding system was small or large.

4.1 Missing Data and Imputation

If only complete cases were used, just 41.4% of survey samples would be retained, potentially reducing the power of the study to unacceptable levels. To handle missing data, there were three potential strategies:

1. **Listwise-deletion:** Remove rows that contain missing data. This would reduce the strength of the dataset.
2. **Mean/median substitution:** Take the mean/median of the existing data points and substitute these values into the missing data points. This would bias the analysis since it decreases variance.
3. **Multiple imputation:** Rather than replacing missing values with a single value, use the distribution of the observed data/variables to estimate multiple possible values for the data points. (See Appendix B for a detailed description of the imputation methodology)

The multiple imputation method was ultimately chosen to fill in missing values, as it accounts for the uncertainty around the true value, and obtains approximately unbiased estimates (under certain conditions). Moreover, this method allows for calculation of standard error around estimations, which in turn leads to a better sense of uncertainty for the analysis.

4.2 Survey Weighting

While a concerted effort was made to obtain a representative sample during the initial collection stage, there was still some bias in the survey sample due to non-response and voluntary responses. In order to accurately estimate statewide numbers, weights were assigned to each system. As described in Valliant et al. (2018), a weighted total estimate has the form

$$t = \sum_{s} w_i \, y_i$$

where t is the total, y_i is a response provided by the ith sample member and w_i is the corresponding analysis weight. Without the use of weights, estimates from survey data would simply reflect nuances of the survey sample, containing the pre-existing survey biases. The procedure for calculating the weights is outlined in Appendix C.

4.3 Margins of Error

The general equation for calculating margin of error is

$$z * \sqrt{\frac{p * (1 - p)}{n}}$$

where z is the z-score, p is the sample proportion of the population, and n is the sample size. The z-score for a 90% confidence interval is 1.645, and the z-score for a 95% confidence interval is 1.96.

4.4 Qualitative Analysis

Each of the two surveys had unique questions, and so the qualitative analysis was conducted separately; however, the methodology used for both was the same. Prior to review, staff created an initial coding schematic. These initial codes were for comments related to: "software limitations," "staff time limitations," "mention of SB998," "desire to shutoff water," "combined billing," "increased cost," etc.; comments were coded according to this schematic. During this review process, staff noted other potential codes that appeared, including specific types of software limitations (inability to search for prior records, inability to separate according to customer type, inability to separate zip codes), as well as other financial impacts, like "delay of capital improvement projects," "increased costs due to COVID (purchase of PPE, increased PTO for staff)," "small systems financially associated with other nearby systems." After these new comment types were noted, all comments were reviewed again to ensure that the review was completed at least twice with the full coding schematic. This process allowed staff to determine findings and summarize common difficulties and concerns from water systems related to COVID-19.

4.5 Financial Vulnerability Assessment (Small Systems Only)

Three separate financial vulnerability metrics (financial risk, revenue to expense ratio, and days of unrestricted cash on hand) were calculated to assess system financial health. Each of the three metrics were assessed as a numerical score from 0 to 5, with 0 representing "very low" risk and 5 representing "extreme" risk. A separate combined vulnerability score was determined based on the results of the three metrics.

4.5.1 Financial Risk

Financial risk was a self-reported multiple choice question in the small system survey that asked respondents, "How many months do you estimate that you will be able to go before needing financial assistance?" Risk was assigned as follows:

- Very low (0): No financial assistance anticipated
- Low (1): More than 12 months
- Medium (2): 9 to 12 months
- Medium High (3): 6 to 9 months
- High (4): 3 to 6 months
- Extreme (5): 0 to 3 months

4.5.2 Revenue to Expense Ratio

Revenue to expense ratio is defined as total reported revenue for the April-October reporting period, divided by total reported expenses for April-October. Risk was assigned as follows:

- Very low (0): Ratio is greater than 2
- Low (1): Ratio is 1.5 to 2
- Medium (2): Ratio is 1.2 to 1.5
- Medium High (3): Ratio is 0.95 to 1.2

- High (4): Ratio is 0.5 to 0.95
- Extreme (5): Ratio is less than 0.5

4.5.3 Days of Unrestricted Cash on Hand

Days of unrestricted cash on hand is defined as the reported unrestricted cash reserve at time survey was taken, divided by estimated daily expenses for report period (April-October). Estimated daily expenses are the total reported expenses for the report period, divided by the number of days from April to October. If only total cash reserve was reported, unrestricted cash was assumed to equal the total; if unrestricted cash was not reported, but restricted cash was reported, then unrestricted cash was assumed to be the difference between the total cash reserve and reported restricted cash. Risk was assigned as follows:

- Very low (0): More than 365 days of unrestricted cash on hand
- Low (1): 180 to 365 days
- Medium (2): 90 to 180 days
- Medium High (3): 60 to 90 days
- High (4): 30 to 60 days
- Extreme (5): less than 30 days

4.5.4 Final Vulnerability Score

The final vulnerability score is the count of "high-risk" (i.e. scores of 4 or 5) per system. The final scores were determined as follows:

- Low: 0 out of 3 high-risk indicators
- Medium: 1 out of 3 high-risk indicators
- High: 2 out of 3 high-risk indicators
- Extreme: 3 out of 3 high-risk indicators

Appendix A Completeness by question type

This section provides an example of the variability in question completion, and how it relates to system size.

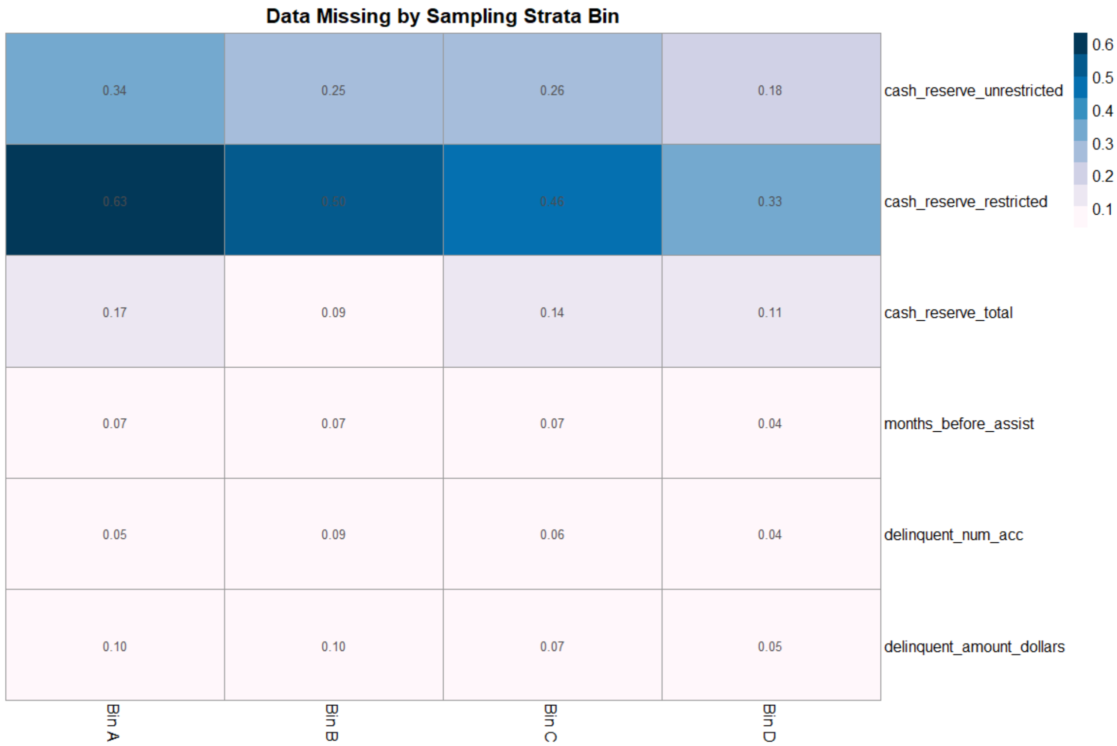

Figure 3: Completeness for some of the small system survey questions and the corresponding survey bins

Figure 3 shows the proportional amount of missing responses for some of the questions used to determine a system's financial vulnerability, with the system size bin on the horizontal axis and the question name on the vertical axis. There seems to be a clear step-wise trend in the response rate for the unrestricted (cash_reserve_unrestricted) and restricted (cash_reserve_restricted) cash reserve variables, with the smallest systems (Bin A) containing a substantial proportion of non-responses (63%) compared to the Bin D systems (33%). In some cases, unrestricted cash reserves are not applicable to small systems that do not have restrictions, so these missing values are not an issue; in other cases, the system did not have finances readily available and were unable to provide a detailed breakdown. Other variables, such as the total amount of debt from delinquent accounts (delinquent_amount_dollars), had a much higher completion rate, with missing responses ranging from 5-10% for the four bins.

Further statistical analysis, as well as qualitative District staff responses, indicated that the completeness of finance reporting was reflective of the system's relative size.

Appendix B Multiple Imputation

The basic steps for imputation of missing data are described below (Katitas, 2019).

1. Estimate the missing values by using an appropriate model which incorporates random variation.
2. Repeat the first step 3-5 times.
3. Perform the desired analysis on each data set by using standard, complete data methods.
4. Average the values of the parameter estimates across the missing value samples in order to obtain a single point estimate.
5. Calculate the standard errors by averaging the squared standard errors of the missing value estimates.
6. Calculate the variance of the missing value parameter across the samples.
7. Combine the two quantities in multiple imputation for missing data to calculate the standard errors.

More simply, the steps of multiple imputation can be described as follows (Kropko et al. 2014):

1. Choose values that keep the relationship in the dataset intact in place of missing values
2. Create independently drawn imputed datasets
3. Calculate new standard errors by measuring the variability across imputed datasets

Appendix B.1 Missing Data Assumptions

Data missingness can be described in three distinct ways (Rubin 1976):

1. Missing Completely at Random (MCAR): Missingness of data points are random, meaning that the pattern of missing values is uncorrelated with the structure of the data.
2. Missing at Random (MAR): Missingness is not completely random; however, the propensity of missingness depends on the observed data, not the missing data. In this case, the missing value can be predicted by looking at the answers for the personal information question.
3. Missing Not at Random (MNAR): Missingness is not random, and correlates with unobservable characteristics

Like all multiple imputation techniques, this effort started with the MAR assumption. While MCAR is desirable, in general it is unrealistic for the data; thus, it was necessary to assume that missing values could be replaced by predictions derived by the observable portion of the dataset. This fundamental assumption allowed for the plausible prediction of missing values.

Appendix B.2 Multiple Imputation Methodology

Conditional Multiple Imputation is utilized to impute missing data from this survey. This approach follows an iterative procedure, modeling the conditional distribution of a certain variable given the other variables. This technique allows for enhanced flexibility as a distribution is assumed for each variable rather than the whole dataset.

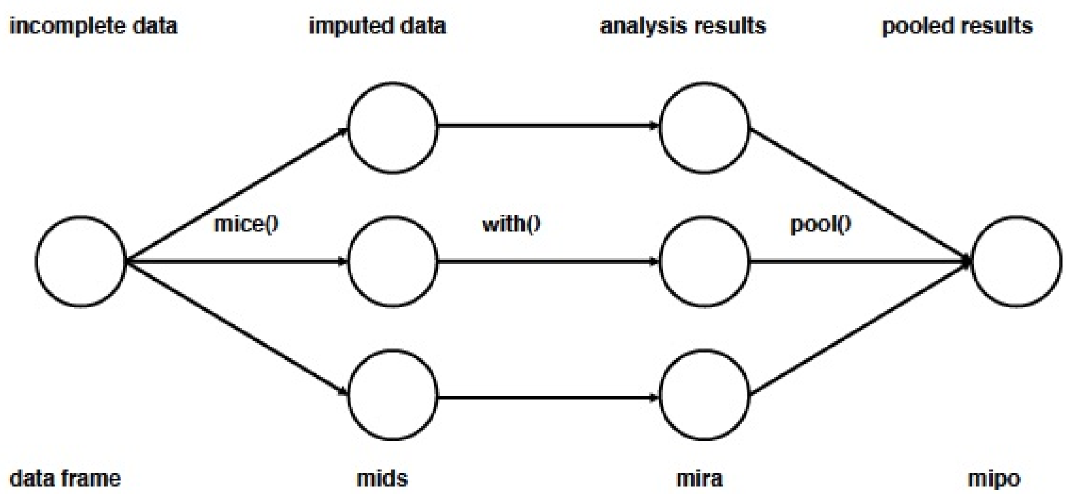

Figure 4: Steps to impute data, using R. Source: University of Virginia Library

The imputation procedure will be demonstrated using cash reserve as an example.

Step 1: Initial Prediction. A statistical model (in this case, random forest modeling) was used to predict cash reserve based on relevant predictors for all systems (e.g. service connections, median 12 month household income, delinquent number of accounts, total delinquent debt, and CalEnviroscreen scores). A significant number of observations were deleted due to missingness (76 out of 371), which is greater than 14% of the whole dataset.

Step 2: Data Transformation. Variables with more than 25% missing values were removed, as well as variables that were highly correlated with others. Additionally, extreme outliers were removed so as to not impact results.

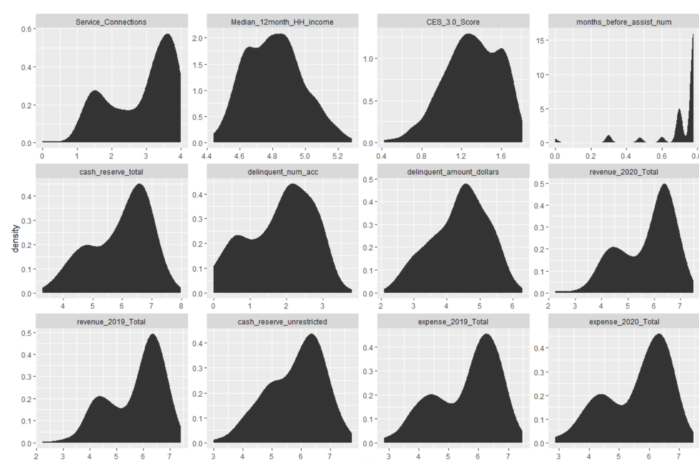

Figure 5: Log-transformed variables that were used in the imputation calculations.

Step 3: Imputation. The mice package assumed a distribution for each input variable, and estimated missing values according to the resultant distributions. Using an iterative Markov Chain Monte Carlo algorithm, the program generated multiple estimates of missing values, which eventually converged to the final values as the iterations progressed.

Step 4: Diagnostics. Imputed values were inspected to ensure that the estimates were plausible values, i.e. values that could have been observed if they had not been missing. Plausibility was checked both graphically for each variable, and statistically by comparing summary statistics and relationship with against the observed values. The diagnostics results determined that the missing data were missing completely at random due to the imputations having the same distributions as the observed data. Therefore, the imputed values were realistic.

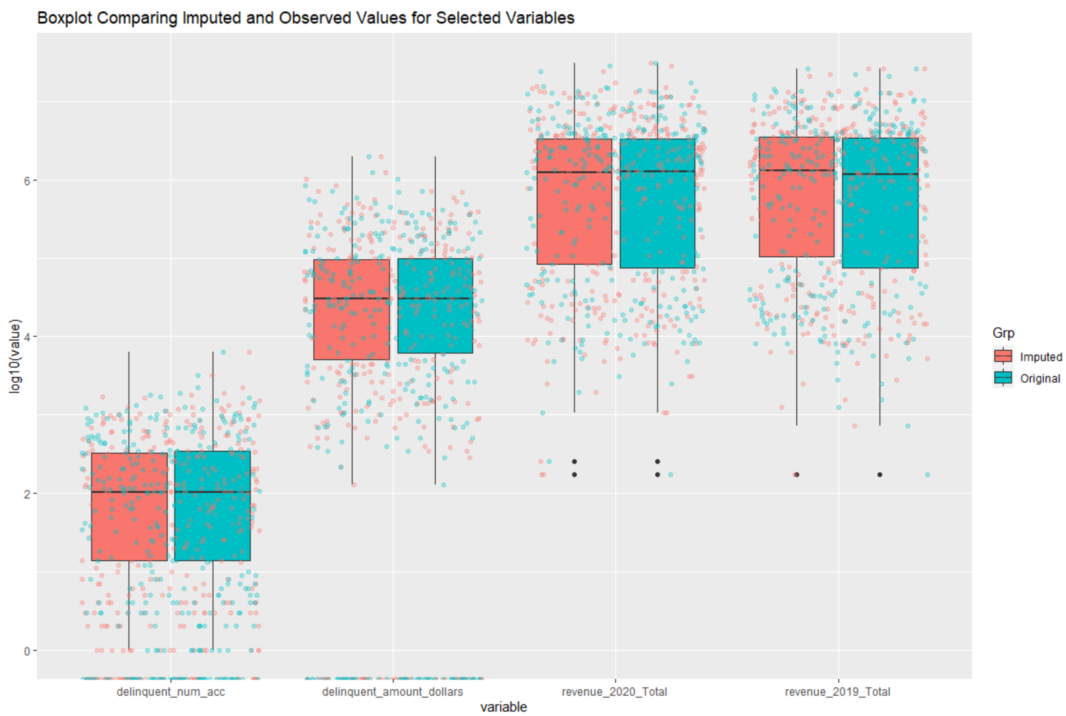

Figure 6: Comparison of selected variables using only original dataset values (blue) versus the dataset with imputed values included (red). The median and quartile values have not changed appreciably.

Appendix C Survey Weight Calculation Methodology

The series of weighting applied to this survey data included computation of base weights, nonresponse adjustments, and use of auxiliary data to reduce variances (i.e. calibration).

Appendix C.1 Base Weights

The first step took the different sampling probabilities of respondents into account, simply based on proportions in the population. This is also known as generating *base weights* (also known as design weights, as used here). Note that this survey was a random stratified sampling design without replacement. Base weights were therefore produced by calculating inclusion probabilities for each strata for the small systems (Bins A, B, C, D). Large systems (>10,000 service connections) were treated as a single strata for the calculation of base weights, and were handled separately.

To correct for these differential probabilities, design weights are calculated so that the sample does not over- or under-represent relevant groups. The design weights are equal to the inverse of the probability of inclusion to the sample. Therefore, the design weight (d_0) of a respondent (i) will be equal to

$$d_0[i] = 1/p_i[i]$$

where p_i is the probability of that unit being included in the sampling.

A simple interpretation of design weights is "the number of units in the population that each unit in the sample represents". **The sum of all design weights should be equal to the total number of units in the population.**

Appendix C.2 Response Propensity Adjustment

After the calculations outlined in the previous section, the weights were further adjusted to account for differences in response rate. It's possible that certain profiles (e.g. lower income communities) had different propensities to respond than another profile (e.g. higher income communities). To ensure that the weights properly reflect the characteristics of the sample as a whole, further adjustments must be made.

Valliant et al (2018) recommend estimating the response propensities and grouping them into classes. Nonrespondents were classified using a classification and regression tree (CART), as shown in Figure 7.

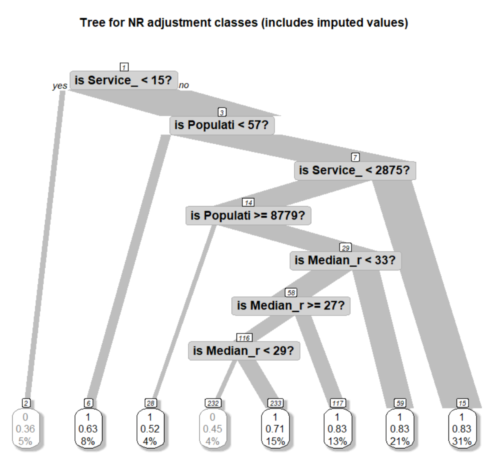

Figure 7: Decision tree classification of non-respondents in the small survey data

The CART estimates were further adjusted using random forest calculations.

Appendix C.3 Weight Calibration

The final step in the weights calculation methodology used auxiliary data (i.e. household income, service connections, population, regulating agency, fee code) to correct coverage problems and reduce standard error, thus improving the survey representativeness. The relationships between the auxiliary variables and the survey variables were modeled using a general regression estimator (GREG) approach, and the resultant model was used to perform the final calibration on the weights.

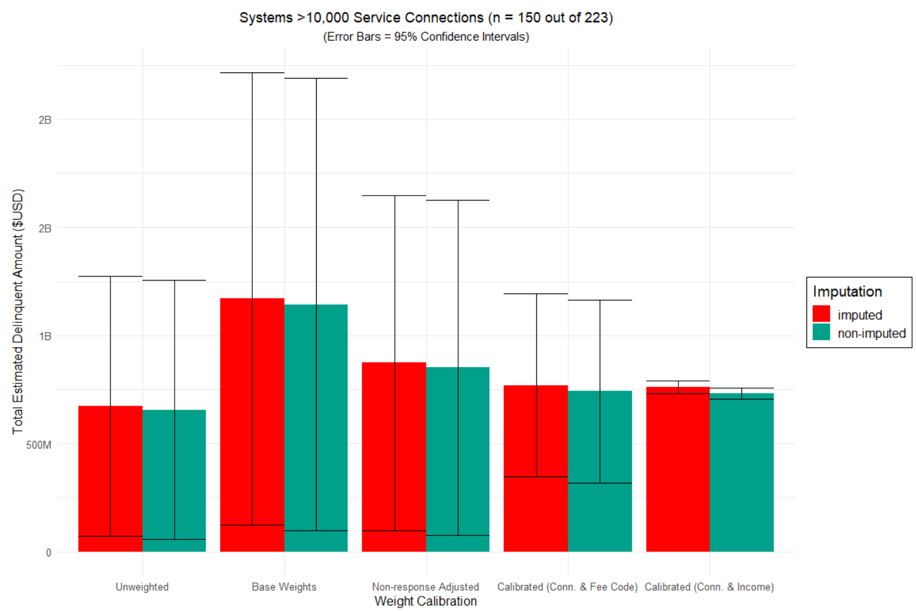

Figure 8: Unweighted, base weighted, non-response adjusted, and calibrated estimates of total debt for the large system survey

Figure 8 demonstrates the impact of using unweighted data versus the weights at each step of the weights calculation process.

Appendix C.4 Variable Importance Using Random Forest

While this is a small dataset and it would be difficult to use this algorithm in a predictive capacity, it is able to provide other information, such as which variables help best predict the months before assistance is needed. A random forest of 10,000 trees and a maximum depth of 5 levels from the open source *scikit-learn* library was fit to all the survey data with the aim of predicting the months before assistance is needed for each system. Table 1 shows the percent importance of each variable for seven different random forest fits. In the first fit, all of the variables are given to the algorithm and the relative importance of the variables are stated as percents. The least important variable was dropped for the next random forest fit, and the results recorded. This was repeated until only one variable remained, the median household income, which had an accuracy of 68.4%.

Median household income and number of service connections were determined to be the most important variables for predicting the months before assistance response. There is a large range for these values, which generally show that assistance is needed sooner for smaller systems with lower median household incomes.

Table 1: Random forest determination of significant variables for predicting months before financial assistance needed

Accuracy	71.1%	72.2%	70.8%	69.1%	66.3%	68.4%	68.4%
Median Household Income (MHI)	18.1%	20.2%	23.3%	30%	35.3%	52.1%	100%
Service Connections	14%	15.1%	20.2%	23.7%	33.5%	47.9%	-
Renter Income %	15.2%	17.2%	19.2%	23.6%	31.2%	-	-
Total Cash Reserves	15.4%	16.7%	19.4%	22.7%	-	-	-
CES Score	14.3%	15.8%	18%	-	-	-	-
Population	13.6%	14.9%	-	-	-	-	-
District (#)	9.5%	-	-	-	-	-	-
Number of Systems Considered	363	363	363	363	377	377	377

Appendix D References

Breiman L. (2001). Random forests. Machine Learning 45:5–32.

Hartley H. O., Rao J. N. K. (1962). Sampling with unequal probabilities and with replacement. Annals of Mathematical Statistics 33(2):350–374.

Katitas, Aycan. 2019."Getting Started with Multiple Imputation in R." University of Virginia Library. Obtained from https://data.library.virginia.edu/getting-started-with-multiple-imputation-in-r/)

Kropko, Jonathan, Ben Goodrich, Andrew Gelman, and Jennifer Hill. 2014. "Multiple imputation for continuous and categorical data: Comparing joint multivariate normal and conditional approaches." Political Analysis 22, no. 4.

Rubin, Donald B. 1976. "Inference and missing data." Biometrika 63, no. 3: 581-592.

Valliant, Richard, Jill A. Dever, and Frauke Kreuter. 2018. Practical Tools for Designing and Weighting Survey Samples. Statistics for Social and Behavioral Sciences. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-93632-1.

Zhao, Peng et al. "Propensity score and proximity matching using random forest." Contemporary clinical trials vol. 47 (2016): 85-92. doi:10.1016/j.cct.2015.12.012