Impact of Certified Mail on Response Rates[1]

Redouane Betrouni, Peter Schilling, Dedrick Owens
U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

Redouane.Betrouni@census.gov, Peter.Schilling@census.gov, Dedrick.L.Owens@census.gov

Abstract

The Annual Survey of Public Employment and Payroll (ASPEP) is conducted by the U.S. Census Bureau to collect data on federal, state, and local governments' civilian employees and their gross payrolls. Quality of survey estimates can decrease in the presence of low response rate due to the effects of nonresponse bias. We conducted an experiment in the 2017 Census of Governments: Employment Component (the government census-year version of ASPEP) to measure the response rates of units receiving Certified mail follow-up reminders versus 1st Class mail follow-up. The study design sought to eliminate factors that might cloud the contrasting impact of competing treatments. In this research, supervised machine learning methods were used to construct certain design features for the experiment and data analysis was conducted using both Bayesian statistical inference and traditional methods. The research concludes that using Certified mail follow-up increases response relative to 1st Class mail follow-up. We recommend Certified mail follow-up as a cost and quality effective option when telephone follow-up cost per unit exceeds twenty-six dollars.

Key Words: Experimental design, Classification regression trees, Bayesian inference, Nonresponse

1. Introduction

Due to the general decline in response rates in government surveys (Czajka et al, 2016), subject matter experts and researchers are currently looking at ways to improve survey response. Some studies show improvements in response through use of advance letters and reminder cards (Scott, 1961). Our research was conducted using the 2017 Census of Governments: Employment Component (CoG-E).

1.1 Census of Governments: Employment Component

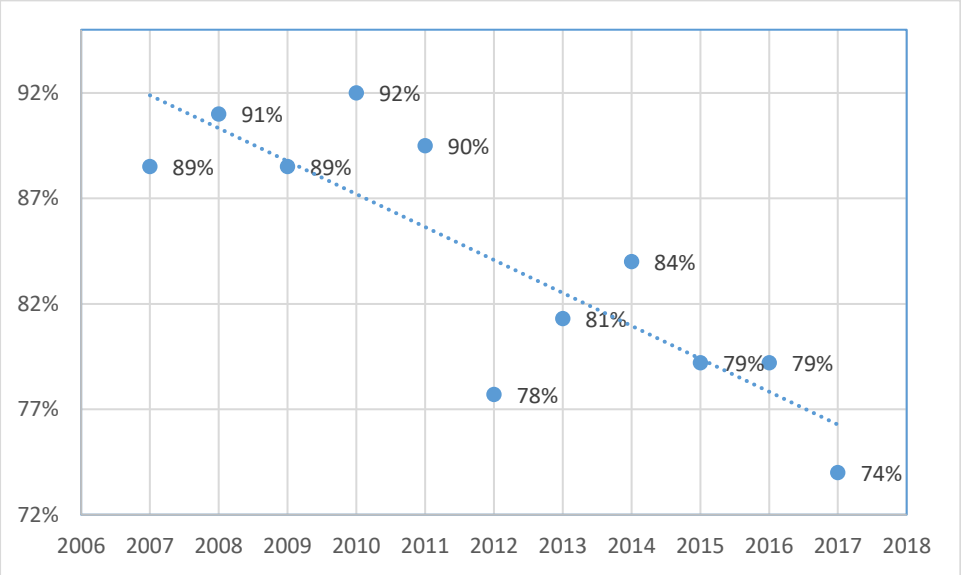
The CoG-E is conducted by the U.S. Census Bureau every five years, in years ending in 2 and 7, to collect data on more than 90,000 state, and local governments' civilian employees and their gross payrolls. In the years in-between the government censuses, the Annual Survey of Public Employment & Payroll (ASPEP) is conducted to collect similar data on a nationally representative sample basis. The CoG-E collects government data classified into five types of governments: counties, cities, townships, special districts, and school districts. The different types of governments perform various governmental activities which are designated for the survey by governmental function codes.

[1] *Disclaimer: Any views expressed are those of the authors and not necessarily those of the U.S. Census Bureau. The Census Bureau has reviewed this data product for unauthorized disclosure of confidential information. (Approval ID: CBDRB-FY19-ESMD-B00014).*

1.2 Historical Response

Quality of survey estimates can decrease in the presence of low response rate due to the effects of nonresponse bias. The scatter plot in Figure 1 shows a downward trend in ASPEP response rates over the last 11 years; with a decline of more than 11 percentage points from the previous year in the 2012 CoG-E year response rate. This survey's response rate has fluctuated from a maximum of 92 percent in 2010 to 74 percent in 2017.

Figure 1*:* Response Rates for Annual Survey of Public Employment & Payroll (ASPEP) from 2007 to 2017 (Note: For 2007, 2012 and 2017 the information is from the CoG-E.)



Source: https://www.census.gov/programs-surveys/apes/technical-documentation/methodology/annual.2017.html

This survey is the only source of public employment data by governmental program function and full-time and part-time work. We thus explored ways to improve response to collection efforts of these critical data.

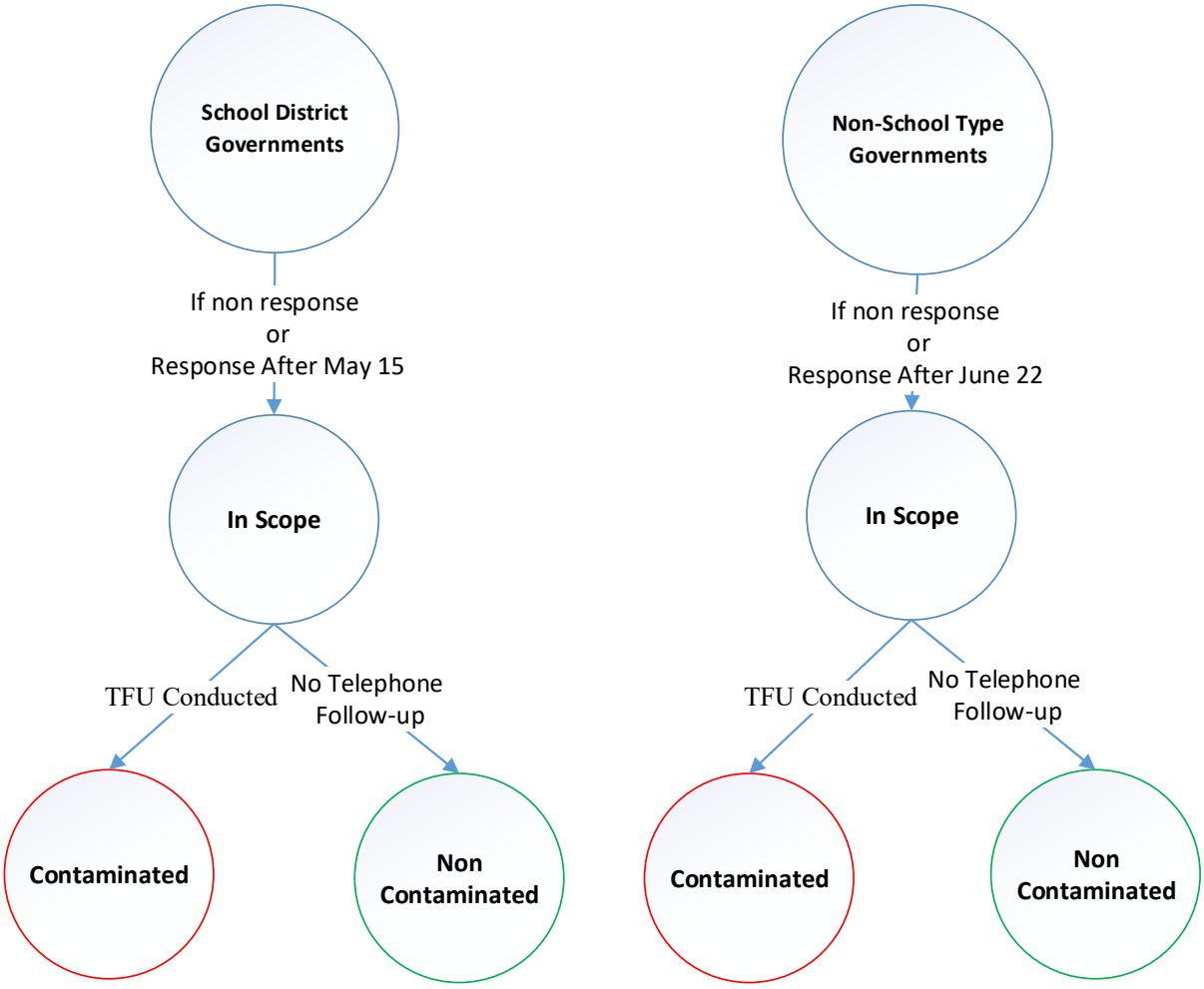
1.3 Study Target Population

It is common practice for a survey contact strategy to consist of multiple attempts to elicit response. The 2017 COG-E featured an introductory letter soliciting online reporting, followed by a pre-due e-mail within 4 weeks, a 1st Past due reminder e-mail 3 weeks after the due date, 2nd Past due reminder letter 2 weeks later, and finally a call from the Telephone Follow-up operation. The experimental focus is the level of response to 2nd follow-up letters in envelopes with green certified stamp versus letters inside standard 1st class mail envelopes. Units responding prior to 2nd Past due reminder or contacted by telephone follow-up post 2nd Past due reminder were deemed out-of-scope.

The target population consists of the units that had not responded to the survey at the time the follow-up started for this research and not yet contacted by the Telephone Follow-up (TFU) operation. This resulted in a total of 29,303 units. Attachment D provides actual timing for data collection activities.

The diagram in Figure 2A reflects the spring 2017 data collection timing of school districts and non-school type governments. It also distinguishes units whose response had been "contaminated" by TFU and thus out of scopes for the study.

Figure 2A: Determination of In-scope Units for the Study on Impact of Certified Mail Follow-up



Numerous studies have been conducted in the past to investigate whether different methods of contact strategies, such as sending post card reminders to non-responders or follow up letters can lead to improved response rates. Marquette et al (2012) tested the effectiveness of multiple contact strategies on improving response during the 2012 economic census. One of the strategies included placing a green sticker on the mailed package to give a more official bearing, which we refer to as the certified mail follow-up strategy, Certified mail follow-up

strategy was tested against the traditional regular mail which we refer to 1st Class letter follow-up in this paper.

Unlike Marquette we tested a single contact and used a subset of the 2017

Based only on one contact strategy, we conducted an experiment using a subset of the 2017 CoG-E nonresponding units (29,303) prior to the mailing of the second reminder letter to determine whether the use of certified letter follow-up improved response rates relative to the 1st Class letter follow-up. We also evaluated the tradeoff between improved response and the increased costs (relative to 1st Class mail) to determine that price point at which Certified mail is the more cost effective option.

2. Experimental Design

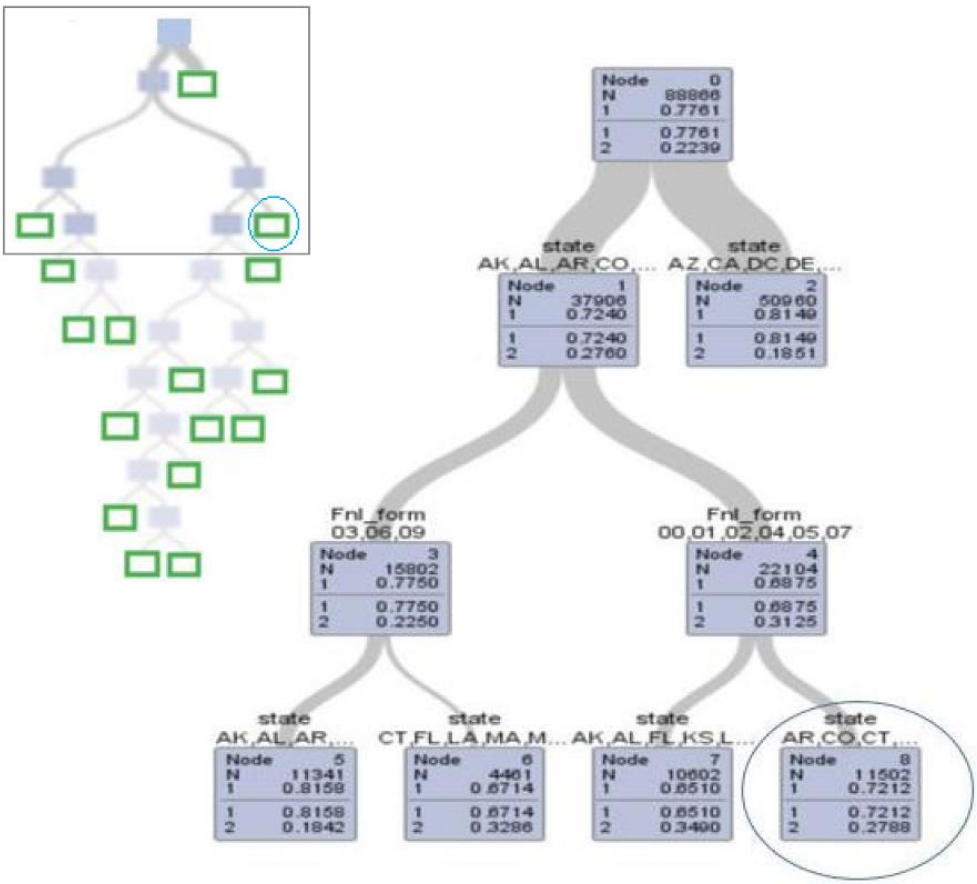
When assessing the effectiveness of the Certified vs. 1st Class mail follow-up on response, it is necessary to rule out alternative explanations —due to variables other than the treatment envelope type (i.e., certified versus 1st class envelope)—for any observed treatment effect. We accomplished this by creating subgroups, referred here as blocks, of similar units with respect to response propensity and assessing whether there is a significant difference in response to these competing treatments as measured within subgroups/blocks.

2.1 Definition of Blocks using Machine Learning

We conducted an empirical study with five years of prior data (2012 to 2016) where a Classification and Regression Tree (CART) application generated a response propensity model and predicted future nonresponse. The empirical study included data from the prior CoG-E in 2012 and the ASPEP sample from years 2013 through 2016. We fitted CART using the 2012-2016 data. This involved training and validating a "Tree" model. The CART model formed homogenous groups (represented as hollow green nodes in the tree silhouette of Figure 2) with respect to response propensity scores—this means that all training set units within final groups had similar response propensity. From an experimental design perspective, CART partitioned the target population into subgroups such that the variability within subgroups is less than the variability between. These subgroups, shown as green "end nodes" in Figure 2B, were designated as blocks for the experiment. Attachment A provides a description of the variables used for the CART model.

This essential blocking feature helps to eliminate the effects of confounding variables—like collection mode, government type, and location— when assessing the relative effectiveness of treatments. Without this blocking it would be difficult to distinguish whether any response differences detected are due to the follow-up letter. For example, regional biases or local government independence may manifest in the form of nonresponse. The blocks help us assess competing follow-up method effectiveness because within block comparisons eliminate geography, collection form, and government effects. When controlling for these extraneous factors, block level test results generally reflect the true impact of the treatment.

Figure 2B. Tree Used to Define Block Groups for Experiment



The silhouetted tree that falls within the rectangular is shown (on the right side) magnified for additional explanation. CART information about the green circled end node, also circled for reference in the magnified sub-tree, are as follows:

- Node 8 as defined by CART (which we now assert to be **Block 8** for our experiment) is described as "all units in one of the states = AR, CO, CT, MN, NH, NM, OH, PA, RI, and VT, FORM Number = 00 (Central Collection), 01(State Agencies), 02(State Inst. Of Higher Ed.), 04(Counties), 05(Townships), 07(Special Districts)".
- There were 11,502 set units falling into this group at the model build stage.
- The historical response rate for units in Node 8 is 27.0 percent.

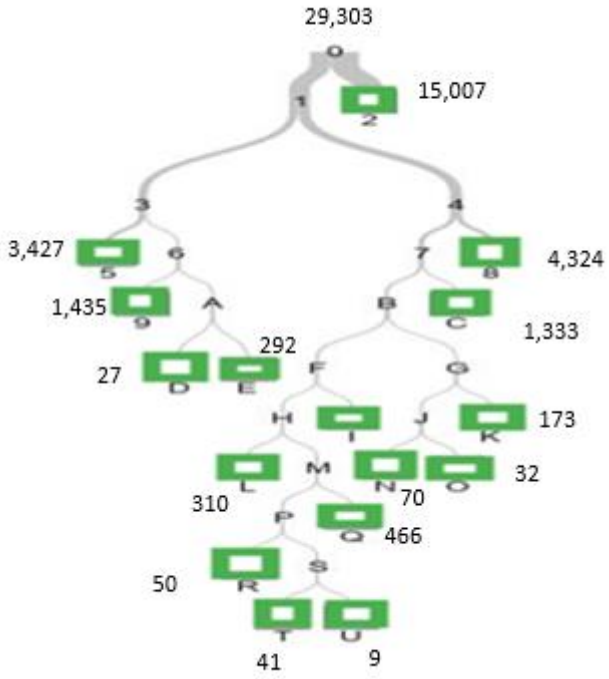
There were 16 block groups (nodes) as defined by CART, corresponding to the tree locations shown in green. Tree node 8 is selected to illustrate tests conducted for each block group and at the national level. Attachment B provides a similar level of detail for the five blocks groups in the silhouetted tree that falls within the rectangle of Figure 2B.

2.2 Assignment of Treatments

The 2017 CoG-E units, approximately 90,000, were assigned to the 16 blocks prior to initial mail-out in accordance with definitions created by the CART model. These units were then randomly assigned, within each block, to receive either Certified or 1st Class mail follow-up. After filtering units that had not responded and had not been contacted by TFU –by the second reminder mail-out--- the target population consists of 29,303 experimental units. Attachment B also reflects the total number of Certified and 1st Class mail designated units assigned to each of the block groups described prior to initial mail-out.

Figure 3 reflects the block assignments of the 29,303 nonresponding units at the start of the 2nd reminder follow-up that were not contacted by TFU. These nonresponding 2017 CoG-E units will be the focus of our investigation into the relative effectiveness of Certified vs 1st Class mail follow-up.

Figure 3. Counts of 2nd Reminder Letter units Assigned to Block Groups.



Further, as an example we now examine response in the wake of the 2nd follow-up reminder for Block group 8. Table 1 reflects a 34 percent response rate for the Certified mail recipients, and 17 percent rate for 1st Class.

Table 1. Block Group 8: Example Data Collection Results

Type of Follow-up	Nonresponse	Response	Observed Response Rate	Total Outstanding at the 2nd Follow-up
Certified mail	1513	771	34%	2,284
1st Class mail	1,703	337	17%	2,040
Total		1108	26%	4,324

It is clear that the observed response rate for Certified mail response is higher than that of 1st Class mail. However, we must determine if this result is statistically significant and consistent across all block groups and at the national level. In the next section, we lay out the methodology by which we shall test these claims.

3. Evaluation of Certified Mail Effect

3.1 Frequentist Hypothesis Testing

A substantial amount of literature is related to statistical inference when comparing two binomial proportions. The classical method of hypothesis testing to draw inference when comparing two proportions is based on the Wald interval. Some methods use profile scores or profile likelihood ratio intervals as improvements.

In our framework, let Y_1 and Y_2 denote the counts of respondents in the Certified mail and 1st Class mail follow-up groups respectively. These two variables can be calculated as

$Y_1 = n_1 - NR_{cert}$ and $Y_2 = n_2 - NR_{1^{st}}$, where n_1 and n_2 are the number of units that were assigned to the certified and 1st Class treatments and $\boldsymbol{NR}_{cert}$ and $\boldsymbol{NR}_{1^{st}}$ are the count of nonrespondents from each group. Y_i follows a binomial distribution for n_i trials with parameter π_i denoted by $Y_i \sim bin(n_i, \pi_i)$. Let $\hat{\pi}_1 = \frac{Y_1}{n_1}$ and $\hat{\pi}_2 = \frac{Y_2}{n_2}$ denote the two response sample proportions for each group. Our goal is to perform statistical inference for the parameter $\pi = \pi_1 - \pi_2$, which is the difference of the two proportions. We assume that the two treatment groups represent independent samples because units were randomly assigned to respective treatments.

We performed the following hypothesis testing:

H_0: $\pi_1 - \pi_2 = 0$ H_a: $\pi_1 - \pi_2 > 0$ at an alpha level of .10, which is a U.S Census Bureau standard of significance requirement for any published report.

Using the Test Statistic given by:

$$T = \frac{\widehat{\pi_1} - \widehat{\pi_2}}{\sqrt{\tilde{\pi}\,(1-\tilde{\pi})\,\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

Under the null hypothesis, π can be estimated using the pooled proportion

$\tilde{\pi} = \frac{Y_1+Y_2}{n_1+n_2}$ given that $\pi_1 = \pi_2$ with the assumption of the equality of the two proportions.

Based on the normal approximation for large sample sizes using the Central Limit Theorem, we can obtain an approximate $100(1-\alpha)$ percent confidence interval for $\pi_1 - \pi_2$ by using:

$$(\widehat{\pi_1} - \widehat{\pi_2}) \pm Z_{\alpha/2}\sqrt{\frac{\widehat{\pi_1}(1-\widehat{\pi_1})}{n_1} + \frac{\widehat{\pi_2}(1-\widehat{\pi_2})}{n_2}}$$

From the 90 percent confidence intervals, we can determine whether the proportion $\pi_1$is statistically significantly greater than π_2 , if the confidence interval of the difference in parameters $\pi_1 - \pi_2$ has a lower bound that is greater than 0 then we will conclude that the Certified mail follow-up treatment significantly increased response rates.

3.2 Bayesian Method

Statistical inference can be made using Bayesian methodology and credible intervals are more intuitive than confidence intervals and simpler to explain. Keeping the same notation as in section 3.1, the number of respondents in the Certified mail follow-up group $Y_1 = n_1 - NR_{cert}$ follows a binominal distribution:

$Y_1 \sim Binomial(n_1, \theta_{cert})$ that has the following density function

$$f(y|n_1, \theta_{cert}) = \binom{n_1}{y} {\theta_{cert}}^y (1-\theta_{cert})^{n_1-y}$$

where θ_{cert} is the hypothetical response rate to Certified Mail at the 2nd reminder follow-up.

Unlike in frequentist statistics, here the unknown parameter θ_{cert} , which is some number between 0 and 1, is considered to be a random variable. We assume that it follows a uniform distribution over the interval [0,1]. This reflects our belief that θ_{cert} can lie in any region between [0,1] such that all subintervals of [0,1] with equal lengths have the same probability.

Bayesian statistics are fundamentally based on Bayes Rule, which allows us to update our beliefs about the θ_{cert} based on new response information Y_1 –essentially a reevaluation of how strongly we believe in values of the parameter given the new information.

For simplicity, in our research we use an uniform distribution on the interval [0,1] as a prior for θ_{cert}, which is a special case of the Beta (α, β) distribution where $\alpha = 1$ and $\beta = 1$. This choice makes the posterior a conjugate posterior, i.e., also a Beta distribution where the parameters are given below.

The assumed uniform prior $\theta_1 \sim U(0,1) = Beta(\alpha = 1, \beta = 1)$ with a density function

$$\mathrm{g}(\theta_1|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} {\theta_{cert}}^{\alpha-1}(1-\theta_{cert})^{\beta-1}$$

results in the joint distribution for(Y_1, θ_{cert}) such that:

$$f(y|n_1, \theta_{cert})\mathrm{g}(\theta_{cert}|\alpha, \beta) = \binom{n_1}{y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} {\theta_{cert}}^{\alpha+y-1}(1-\theta_{cert})^{\beta+n_1-y-1}$$

Hence the posterior distribution for $\theta_{cert} \sim Beta(\alpha^*, \beta^*)$. Let y_1 be the observed value of the random variable Y_1 then $\alpha^* = \alpha + y_1$ and $\beta^* = \beta + n_1 - y_1$

With the following density:

$$g(\theta_{cert}|y_1, \alpha, \beta) = \frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} {\theta_{cert}}^{\alpha^*-1}(1-\theta_{cert})^{\beta^*-1}. \quad (4)$$

Similarly the posterior distribution for $\theta_{1^{st}}$ can be updated as a $Beta(\alpha^*, \beta^*)$ with

$\alpha^* = \alpha + y_2$ and $\beta^* = \beta + n_2 - y_2$ where y_2 is the observed value of the random variable Y_2.

With the following density:

$$g(\theta_{1^{st}}|y_2, \alpha, \beta) = \frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} {\theta_{1^{st}}}^{\alpha^*-1}(1-\theta_{1^{st}})^{\beta^*-1} \quad (5)$$

From the posterior distributions of θ_i, we can construct 90 percent credible intervals. We choose the intervals such that the probability of being below the interval is equal to the probability of being above it, a resulting credible interval using this method is known as equal-tailed interval (ETI).

Obtaining an analytical expression for the posterior distribution for the difference of response rates $\theta_{cert} - \theta_{1^{st}}$ can be tedious; fortunately, a Monte Carlo method can be used to derive the distribution of functions of other random variables. In our problem, direct Monte Carlo simulation can be used because the joint distribution of the pairs (θ_{cert}, $\theta_{1^{st}}$) is the product of the posterior distributions due to the independence.

From the posterior distribution of $\theta_{cert} - \theta_{1^{st}}$ we can obtain the 5th and 95th percentiles, these will be used to determine the boundaries of the 90 percent equal tail credible intervals. If 0 is not included in the interval we will conclude that the test statistic is significant at the alpha level of .10 and we can reject the null hypothesis.

4. Data Analysis Results

4.1 Traditional Data Analysis

4.1.1 Traditional Analysis at the Block Level

Hypothesis testing and confidence intervals were computed using the traditional method based on the Student T test described in section 3.1 for the 16 blocks created by CART.

Table 2A: Experiment Response Counts by Type of Follow-up for 2017 CoG-E by Block

Block	Number of Units	Letter Type	$\hat{\theta}$ (Percent)	$se(\hat{\theta})$	90% Confidence Interval (Percent)	Test $\theta_{diff} > 0$
2	7,119	Certified	35.8	0.006	(34.8 , 36.7)	--
2	7,888	1st Class	20.8	0.005	(20.1 , 21.6)	--
$\theta_{diff} = \theta_{cert} - \theta_{1^{st}}$					(13.7 , 16.1)	Reject H_0
5	1,514	Certified	37.5	0.012	(34.8, 36.7)	--
5	1,913	1st Class	21.4	0.009	(19.8, 22.9)	--

$\theta_{diff} = \theta_{cert} - \theta_{1^{st}}$					(13.6 , 18.7)	Reject H_0
8	2,284	Certified	33.8	0.010	(32.1 , 35.4)	--
8	2.040	1st Class	16.5	0.008	(15.2 , 17.9)	--
$\theta_{diff} = \theta_{cert} - \theta_{1^{st}}$					(15.1 , 19.4)	Reject H_0
9	720	Certified	26.5	0.016	(23.8 , 29.2)	--
9	715	1st Class	13.8	0.013	(11.7 , 16.0)	--
$\theta_{diff} = \theta_{cert} - \theta_{1^{st}}$					(9.2 , 16.1)	Reject H_0
R	21	Certified	28.6	0.099	(12.4 , 44.8)	--
R	29	1st Class	13.8	0.064	(3.3 , 24.3)	--
$\theta_{diff} = \theta_{cert} - \theta_{1^{st}}$					(-4.6 , 34.1)	Fail to Reject

As we can see in Table 2A, there is evidence to conclude that the response rate for the Certified mail group is greater than that of the 1st Class mail group for block level data. There are some blocks where the test failed to reject the Null hypothesis because those are blocks with small counts. This makes the test unreliable; however, the Certified mail group always have marginal response rate higher than those of 1st Class mail group. The lower bounds of the 90 percent confidence intervals were greater than 0, therefore we conclude that the Certified mail follow-up treatment significantly increased response rates. Table 2A displays only 5 blocks and for the few blocks like Block R there were not enough data and the test was not statistically significant but still shows that the response rate was 28.6 percent for Certified and 13.8 percent for 1st Class mail follow-up.

4.1.2 Traditional Analysis at the National Level

Similarly hypothesis testing and confidence interval based on the Student T test to compare two proportions was conducted at the national level.

Table 2B: Experiment Response Counts by Type of Follow-up for 2017 Census of Governments: Employment Component

Type of Follow-up Letter	Nonresponses	Responses	Response Rate (Percent)	Total Outstanding as of mailing of Certified and 1st Class Letter
Certified mail group	9,640	4,908	33.7	14,548
Control group (1st Class mail)	12,022	2,733	18.5	14,755
Total	21,662	7,641		29,303

At the national level, the difference in response rate is closer to 15 percentage points or $\hat{\pi}_{cert} - \hat{\pi}_{1^{st}} \approx 0.337 - 0.185 = 0.152$. This result shows that units in the Certified mail follow-up group are more likely to respond than the 1st Class mail group.

There is statistical evidence to conclude that the response proportion for the Certified mail group is greater than that of the 1st Class mail group at the nation level, the value of the Test statistic is 29.66 and the P-value ≈ 0 , which is the right tail probability from the standard normal distribution that is above the observed value. In fact, we estimate that the actual difference of proportions is between 0.143 and 0.162 (90 percent CI). Therefore, we conclude that the Certified mail follow-up treatment significantly increased response rates by almost 15.2 percentage points. Note that the increased response rates obtained from nonresponding units is due to the certified-mail process.

4.2 Bayesian Test Results

4.2.1 Bayesian Analysis at the Block Level

The following paragraph reflects Bayesian analysis run on our example block (8). Note that this process was repeated for each remaining block and at the national level.

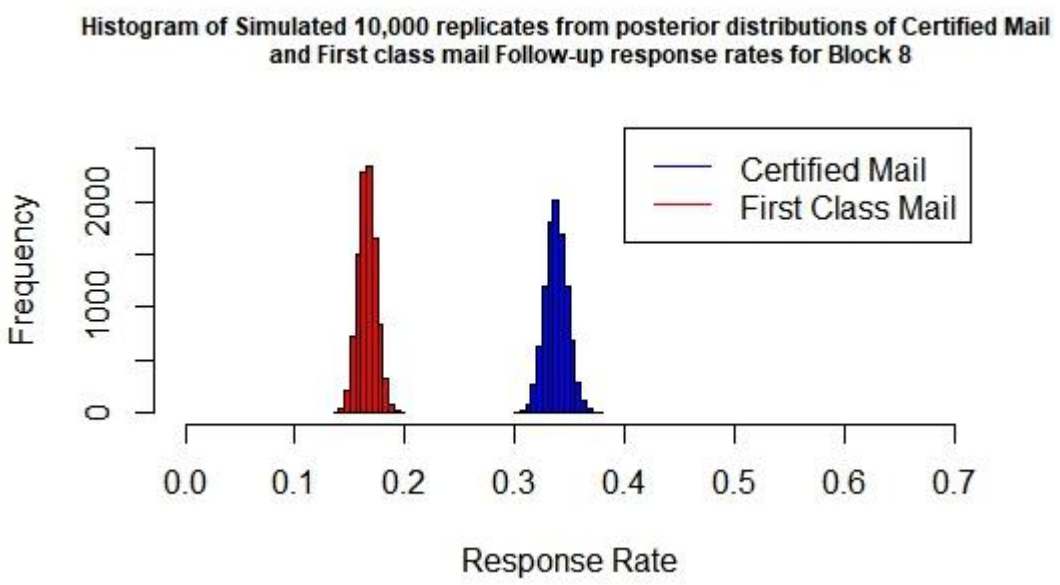
Using equations (4) and (5) the posterior distributions can be updated, and hence the Beta (1 + 711 , 1+ 1,513) and Beta (1 + 337 ,1+ 1,703) are the Bayesian posterior distributions for response rates of the Certified letter and 1st Class letter follow-up groups, respectively.

We used the Monte Carlo simulation method to simulate 10,000 replicates of the joint distribution of the response rates pairs (θ_{cert}, $\theta_{1^{st}}$) for all the Blocks and at the national level according to the posterior Beta distributions obtained above. Attachment C explains why the sample size is set to 10,000 in this experiment as it is derived using power analysis.

As an illustration, the plot for Block 8 in Figure 4 depicts the distribution of certified letter response rates in blue and the distribution of 1st Class letter rates in red.

Figure 4: Histograms of Certified Letter and 1st Class Letter Follow-up Posterior Distribution

Response Rate(s) for Block 8



The distributions in Figure 4 show the certified letter response rate falling between .3 and .4 percent, and the 1st Class mail letter rate between .1 and .2 percent. Note that units at this phase of collection are considered "hardcore" nonrespondents, having been exposed to three or more collection contacts. Their reluctance to respond is reflected in the observed low response rates as shown in the green Table 1 cells and the distribution of low response rates in blue and red in Figure 4.

Figure 5 shows the histogram of a randomly simulated 10,000 replicates from posterior distribution of response rate differences between Certified Mail and 1st Class mail for Block 8. This demonstrates the relative effectiveness of the two follow-up letter treatments, represented by the distribution of the difference in nonresponse rates, where

$$diff = (\theta_{cert} - \theta_{1^{st}}).$$

Note that all the difference values are greater than 0, meaning that in all instances in the observed data the θ_{cert} is greater than $\theta_{1^{st}}$.

Figure 5: Distribution of the Posterior Distribution Difference in Certified Mail and 1st Class Mail Response Rate for 2017 Census of Governments: Employment Component

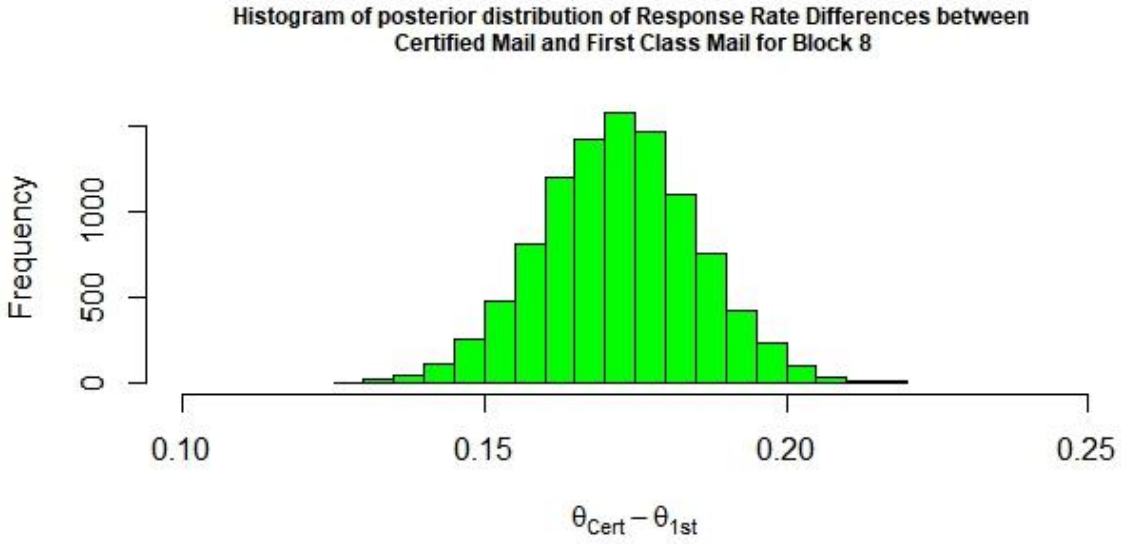


Figure 5 shows that the range of posterior differences $\theta_{cert} - \theta_{1^{st}}$ or θ_{diff}, is between approximately 0.13 to 0.21 percentage points for the 10,000 replicates. We conclude that it is highly probable that θ_{diff} is greater than 0, and this leads us to conclude that θ_{cert} (the certified letter response rate) is larger than $\theta_{1^{st}}$ (the 1st Class response rate). In fact, out of the more than 10,000 rate differences used to construct this plot there were no instances in which the certified rate was less than the 1st Class rate.

Table 3A: Response Rates and Test Results by Type of Follow-up Letter for the 2017 Census of Governments: Employment Data (Block Group 8)

Type of Follow-up Letter	**Nonresponse**	**Responses**	**Response Rate (Percent)**	**90% Equal Tail Credible Interval (Percent)**	**Test** $\theta_{cert} - \theta_{1^{st}} > 0$
Certified	1,513	771	34.0	(32.1 , 35.4)	--
1st Class Mail	1,703	337	17.0	(15.2 , 18.0)	--
$\theta_{cert} - \theta_{1^{st}}$				(15.1 , 19.3)	Reject H_0

Similar analysis was performed on all remaining blocks and at the national level. We found that for all blocks except few which had small sizes the lower bounds of the 90 percent credible intervals were greater than 0, therefore we conclude that the Certified mail follow-up treatment outperforms 1st Class mail follow-up and significantly increased response rates in almost all blocks.

4.2.2 Bayesian Analysis at the National Level

Table 3B shows estimated response rate at the national level by follow-up letter type. The estimate of the response rate to Certified mail follow-up is 0.337; that is, roughly 34 percent of units respond when follow-up is by Certified mail. Based on our research, there is a 90 percent probability that the true rate of response for the Certified mail group is within the interval (0.331, 0.344). Similarly, the response rate of 1st Class mail follow-up is 0.185; i.e., –implying that 18.5 percent of units respond when follow-up is by 1st Class mail. Based on our research, there is 90 percent probability that the true rate of response with 1st Class mail follow-up lies within (0.180, 0.191).

Table 3B: Response Rates and Test Results by Type of Follow-up Letter for the 2017 Census of Governments: Employment Data (National Level)

Type of Follow-up Letter	**Nonresponse**	**Response**	**Response Rate (Percent)**	**90% Equal Tail Credible Interval (Percent)**	**Test** $\theta_{\text{cert}} - \theta_{1st} > 0$
Certified	9,640	4,908	33.7	(33.1 , 34.4)	--
1st Class Mail	12,022	2,733	18.5	(18.0 , 19.1)	--
$\theta_{\text{cert}} - \theta_{1st}$				(14.4 , 16.0)	Reject H_0

A test of whether $\theta_{diff} > 0$, specifically whether zero is not in the credible intervals, leads us to reject the null hypothesis. At the national level we believe there is a 90 percent probability that the increase in response rate θ_{diff} is between 14.4 and 16.0 if the Certified letter is used instead of the 1st Class as a follow-up strategy.

5. Cost Analysis

In this section, we consider the monetary impact of using the certified letter follow-up versus 1st Class letter follow-up. The cost of follow-up via 1st Class letter is \$1.00 and \$5.00 to \$6.00 for certified letter. If units do not respond to these follow-up attempts, as a last resort they are sent to telephone follow-up (TFU). Some calls result in just reminders with average length of approximately 2 minutes, while other calls involve data collection and range from 10 to 30 minutes. Certain surveys have a dedicated unit that is paid regardless of call volume; others pay by call duration or number of completed calls. With these factors in mind, the TFU cost per unit varies from \$10.00 to \$30.00.

A higher certified letter conversion to response at the nonresponse follow-up phase would lead to fewer units being sent to TFU. However, due to higher mailing costs associated with certified letter follow-up, the decision to use this mode instead of the 1st Class letter must take into account mail-out cost ~~as~~ compared to TFU costs.

The expected cost of data collection for each follow-up reminder can be computed as follows:

$E(Cost|Certified) = \$5.5 * n + n * (1 - \theta_{cert}) * tfu_{cost}$ for Certified mail, and

$E(Cost|First) = \$1 * n + n * (1 - \theta_{1^{st}}) * tfu_{cost}$ for 1st Class mail

where n represents the number of units that had not responded by the outset of the 2nd reminder letter mail-out.

The θ_{cert} and $\theta_{1^{st}}$ are the rate of response for certified and 1st Class follow-up reminder respectively.

The difference of costs between using certified versus 1st Class mail follow-up is

$$diff = n\left(\$4.5 - tfu_{cost} \times (\theta_{cert} - \theta_{1^{st}})\right)$$

This cost difference can be estimated as:

$$\widehat{diff}(tfu_{cost}) = n\left(\$4.5 - tfu_{cost} \times \left(\widehat{\theta_{cert}} - \widehat{\theta_{1^{st}}}\right)\right)$$

Certified mail follow-up will be more cost effective if and only if

$$\widehat{diff}(tfu_{cost}) < 0 \qquad (6)$$

Solving for inequality (6), using the posterior modes, which were approximately 0.342 and 0.168 for certified and 1st class mail respectively ,leads to $tfu_{cost} > \$25.83$. This means that using the certified letter for follow-up presents the more cost effective option for a survey (regardless of follow-up workload) when the estimated per unit TFU cost for the survey is more than $25.83.

Figure 6 below gives a broader perspective on the impact of the follow-up phase workload and average TFU cost/unit on the decision to use a certified letter vs. 1st Class letter. The plot uses a progressive color scheme to show the impact on follow-up letter type preference of different workload and TFU cost/unit combinations.

We analyzed the difference of costs between using certified versus 1st Class mail follow-up as a function of TFU cost per unit and work load and partitioned its range into five equidistant cost ranges.

We color coded these five cost ranges using orange, red, light purple, dark purple and blue in order, for example blue is the group that has the lowest cost values for Certified mail if used as a strategy of collection instead of 1st Class mail.

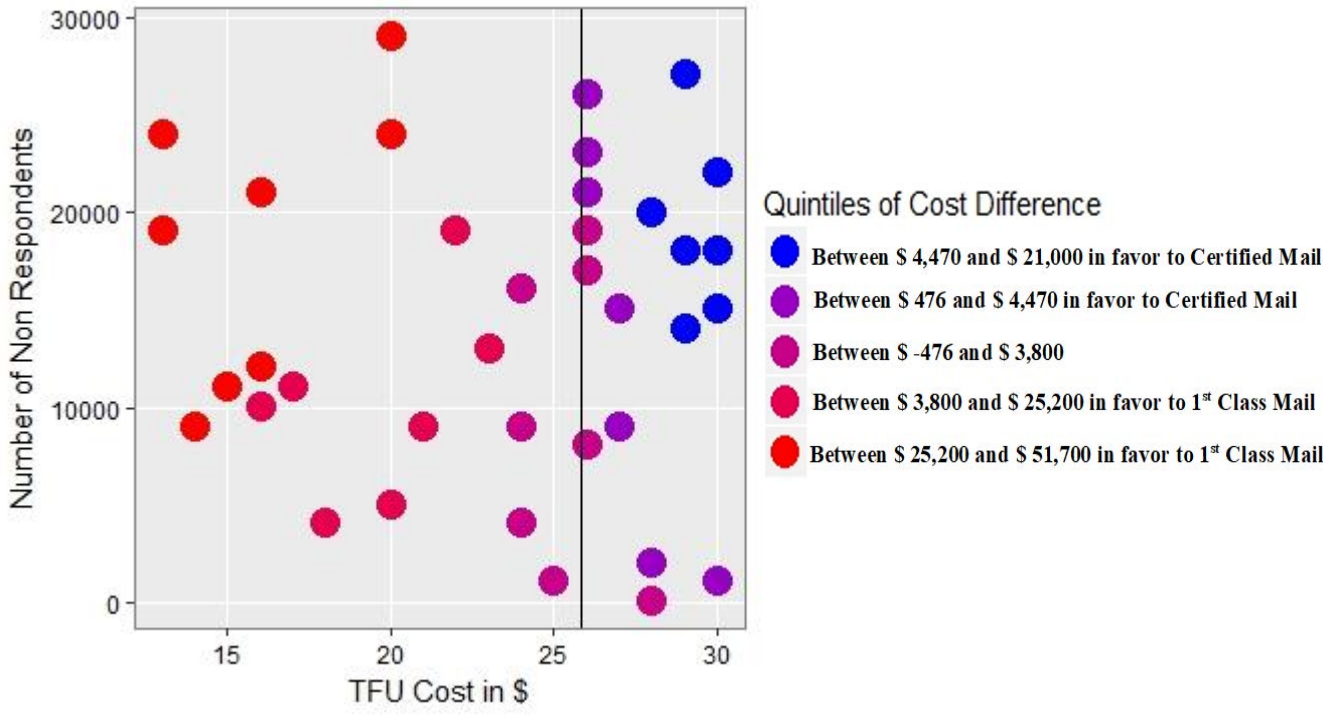
The color coding is summarized in the legend below:

- Any value difference of costs that ranges between $ 25,200 and $ 51,700 in favor to 1st Class mail is coded orange.
- Any value difference of costs that ranges between $ 3,800 and $ 25,200 in favor to 1st Class mail is coded red.
- Any value difference of costs that ranges between $ (476) and $ 3,800 was coded light purple.
- Any value difference of costs that ranges between $ 476 and $ 4,470 in favor to Certified mail is coded purple.
- Any value difference of costs that ranges between $ 4,470 and $21,000 in favor to Certified mail is coded blue.

The X-axis in Figure 6 represents all possible values of TFU cost while the Y-axis represents all possible values of number of nonrespondents before either strategy (treatment) is implemented.

A number of randomly selected points (circles) were selected and depending on their location on the (X,Y) plane their corresponding color is shown instead of their actual value being displayed to demonstrate regions where Certified mail outperforms 1st Class mail in terms of cost and vice-versa.

Figure 6: Certified vs. 1st Class Follow-up Letter Preference Dictated by Mail Follow-up (FU) Workload and Telephone FU (TFU) Unit Cost



The plot reflects the shifting support for 1st Class letter vs. certified letter at differing mail follow-up workload and TFU cost/unit levels. The legend depicts level of support by gradual changes in color as well as changes in cost difference.

With higher TFU cost/unit and smaller workloads, the letter options have similar costs. As workload increases and TFU cost/unit remain high, using certified letter follow-up becomes more cost effective (purple to blue). Given lower TFU cost/unit and smaller workload, the 1st Class letter follow-up is more cost-effective (dark red); this remains true at lower TFU cost/unit levels with increasing workload (dark red to bright red). The vertical line represents the workload and TFU cost/unit combinations where certified and 1st Class letter costs are the same. To make a quick assessment, when TFU cost/unit is greater than $25.83 (right of the vertical line) the certified letter is most cost-effective at any workload level. Combinations to the left of the vertical line tend to favor the 1st Class letter. Therefore, using the certified letter for follow-up presents the more cost effective option for a survey (regardless of follow-up workload) when the estimated per unit TFU cost for the survey is more than $25.83.

6. Conclusion

We conclude that using Certified mail follow-up increases response relative to 1st Class mail follow-up. We recommend that Public Sector surveys follow-up procedures incorporate some

level of Certified mail with proper consideration given to Telephone follow-up costs associated with nonresponding units.

Ultimately, we want to improve quality by increasing response rate and to save costs if the treatment is applied appropriately, that is where the cost-quality threshold points us to the use of Certified mail for follow-up. We determined that overall the use of Certified mail follow-up is cost effective when the TFU per unit cost exceeds $25.83.

We initiated this inquiry with unscientific knowledge of the impact of using certified letters for follow-up in Public Sector surveys. We now know that the certified letter encourages greater response relative to that of the 1st Class letter. However, the decision to use this follow-up mode must be made with due consideration of the follow-up workload and TFU per unit costs. In this study, we have affirmed the benefit of certified letter follow-up, and provided guidance for deciding when its usage is most cost-effective.

7. Future Research

In future research we would like to use an appropriate prior distribution rather than flat prior, where history of response rates is reflected in the prior distribution for the different blocks used in this analysis. We would like to perform sensitivity analysis to determine how sensitive results are to the choice of different choices of prior distribution.

In this research response rate was predicted using CART method, however we can reinforce this binary prediction modeling using additional machine learning algorithms, these include the use of support vector machines, artificial neural networks, *k*-nearest neighbors, regression and Random Forest. Using one or more methods on an ensemble learning algorithm combining the different base models, may increase response prediction accuracy to create better blocks.

Acknowledgments

The authors express gratitude to Magdalena Ramos and Justin Smith for their careful review and thoughtful comments on earlier versions of this report.

References

Agresti, A., & Caffo, B. (2000). "Simple and Effective Confidence Intervals for Proportions and Differences of Proportions Result from Adding Two Successes and Two Failures." The American Statistician, 54(4), 280-288.

Becker, H., Cookston, J., & Kulberg, V. (2000). "Mailed survey follow-ups—Are Postcard Reminders more Cost-Effective than Second Questionnaires" Western Journal of Nursing Research, 22(5), 642-647.

Betrouni, R., Owens, D., Tran, B., & Schilling, P. (2017) "Impact of Certified Mail on Nonresponse Rates" ESMD Internal Memo.

Czajka, J. L., & Beyler, A. (2016). "Declining Response Rates in Federal Surveys: Trends and Implications" (Background Paper). Mathematica Policy Research.

Hollander, Myles, Douglas A. Wolfe, and Eric Chicken. *Nonparametric statistical methods*. Vol. 751. John Wiley & Sons, 2013.

Marquette, E., Kornbau, M., & Toribio, J. (2015, August). Testing Contact Strategies to Improve Response in the 2012 Economic Census. In Proceedings of the Section on Government Statistics: American Statistical Association.

National Research Council. (2013). "Nonresponse in Social Science Surveys: A Research Agenda." National Academies Press.Scott, C. (1961). Research on mail surveys. Journal of the Royal Statistical Society. Series A (General), 143-205.

Attachment A

List of Independent Variables Used for CART Model

- State
- Number of full-time equivalent employees
- Number of full time employees who have worked part-time hours
- Number of full-time employees
- Number of part time employees
- Total full-time pay
- Government type
- Collection Form

Attachment B

Partial List of Classification and Regression Tree (CART) Model Block Specification

Block	Description	2017 CoG-E Units Assigned to		Historical Response Rate
		Certified Mail Panel	1st Class Mail Panel	
2	(***State*** IS ONE OF AZ, CA, DC, DE, GA, HI, IA, ID, IL, IN, KY, MD, ME, MI, MO, MT, NC, NV, NY, OK, OR, SC, TN, TX, UT, VA, WA, WI, WY)	7119	7888	18.0%
5	(***State*** IS ONE OF AK, AL, AR, CO, KS, MN, NE, NJ, OH, PA, SD, WV) AND (***Form*** IS ONE OF 03, 06, 09)	1514	1913	18.0%
8	(***State*** IS ONE OF AR, CO, CT, MN, NH, NM, OH, PA, RI, VT) AND (***Form*** IS ONE OF 00, 01, 02, 04, 05, 07)	2284	2040	27.0%
9	(***State*** IS ONE OF CT, FL, LA, MA, MS, ND, NH, RI, VT) AND (***Form*** IS ONE OF 03, 06, 09)	720	715	29.5%
C	(**Government Type** IS 2) AND (**State** IS ONE OF AK, AL, FL, KS, LA, MA, MS, ND, NE, NJ, SD, WV) AND MISSING (***Form***) **OR** (***Form*** IS ONE OF 00, 01, 02, 04, 05, 07) AND (***State*** IS ONE OF AK, AL, AR, CO, CT, FL, KS, LA, MA, MN, MS, ND, NE, NH, NJ, NM, OH, PA, RI, SD, VT, WV)	718	615	28.5%

Note on Form Number variable values:

FORM Number = 00 (Central Collection), 01(State Agencies), 02(State Inst. Of Higher Ed.), 04(Counties), 05(Townships), 07(Special Districts)

Attachment C

Determination of Monte Carlo Simulation Sample Size

Repeated independent random samples were generated from the posterior distributions and the required sample size was determined using the power analysis described in the next paragraph.

The required number of simulated random samples for Monte Carlo simulations to test a given hypothesis h can be derived using power analysis. If we let p to be the true proportion of errors of the hypothesis h then p can be estimated by assuming the draws to be independent Bernoulli random variables. If we set the goal to have a two-sided 95 percent confidence interval for p with margin of error (MoE) of 0.01. Then the minimum required sample size can be determined.

$$MoE = z_{\alpha/2} se(\hat{p}) = z_{0.025}\sqrt{\frac{p(1-p)}{M}} \approx 2\sqrt{\frac{p(1-p)}{M}}$$

Since $p \in [0,1]$ the quantity $p(1-p)$ has its maximum when $p = \frac{1}{2}$

This can be verified by setting the first derivative to 0, solving for p and verifying that it is a maximum because the sign of the second derivative is negative.

Therefore $MoE \leq \frac{1}{\sqrt{M}}$ and the goal is to have $MoE \leq 0.01$.

So it is sufficient to require $M \geq 10{,}000$.

The Gamma function is defined as follows:

$\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx$, where $\Gamma(0) = 1$.

In particular the gamma function has the following property:

$\Gamma(n) = n!$ for any natural number n

and

$\Gamma(a+1) = a\Gamma(a)$

As a result, the simulations sample size for the experiment is set to 10,000.

Attachment D

E6/E8 Schools Collection Timing

Date	Activity
3/9/2017	ESMD PSSDEB Obtain Initial Mail File/Identify E6 & E8 Schools
3/16/2017	Initial Mail Letter (Non-Certified) to Schools
4/20/2017	Pre-due Reminder E-mail
4/27/2017	COG ASPEP Due Date
4/28/2017	ESMD Provide Initial Mail File E6/E8 Schools with Certified/Non-Certified Assignment
5/1/2017	1st Past due E-mail File Created—reflects removal of responded units
5/1/2017	1st Past due Reminder (Follow-up) File to NPC
5/2/2017	Past due Reminder (Follow-up) E-mail
5/2/2017	Past due Reminder (Follow-up) E-mail
5/7/2017	Updated 1st Past due Reminder file to NPC—reflects removal of most recent respondents
5/15/2017	Mail 1st Past due Reminder (Follow-up)---Certified or Non-Certified Letter
5/15-6/2/17	Targeted TFU for Schools
6/3/2017	Retrieve Post-TFU response data
6/4—9/1/17	ESMD Compile, Analyze, and Report Result

Where ESMD=Economic Statistical Methods division; PSSDED=Public Sectors Surveys Design and Estimation Branch; NPC=National Processing Center.

Non-Schools Collection Timing

Date	Activities
3/16/2017	Initial Mail (Non-Certified)
4/20/2017	Pre-Due Reminder E-mail
4/27/2017	Due Date
5/15/2017	1st Past Due Reminder (Follow-up) Letter
6/08/2017	ESMD Provide File with Certified/Non-Certified Assignment
6/12/2017	2nd Follow-up file to NPC—reflects removal of responded units
6/15/2017	Updated 2nd Follow-up File to NPC
6/22/2017	Mail 2nd Past due Reminder (Certified/Non-Certified) Letter
7/14/2017	Retrieve Response data immediately prior to Past due (Follow-up) E-mails
7/15/2017	Past Due Reminder E-mails
7/15/2017	ESMD Compute, Analyze, and Report Results (Partial)
7/18—9/29/17	Non-Schools Targeted TFU
10/2/2017--*	ESMD retrieve Response data after TFU Close-out---conduct additional analysis as needed