An Embedded Experiment for Targeted Nonresponse Follow-Up in Establishment Surveys

Stephen J. Kaputa[1], Katherine Jenny Thompson[1], Jennifer L. Beck[2]

[1]Economic Statistical Methods Division, U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

[2]National Center for Science and Engineering Statistics, National Science Foundation, 4201 Wilson Boulevard, Arlington, VA 22230

Abstract
The U.S. Census Bureau is investigating adaptive nonresponse follow-up (NRFU) strategies for single unit businesses in the 2017 Economic Census. This paper describes an embedded split-panel field experiment in the 2015 Annual Survey of Manufactures that tests two adaptive NRFU designs. With the first design, nonresponding establishments in the experimental group received a reminder letter either by certified mail (expensive) or standard mail (inexpensive) based on an optimal allocation that assigns a higher proportion of the certified letters to domains that initially have low unit response rates. This targeted allocation procedure ensures that all units receive some form of NRFU, but saves cost over the current procedure that sends a certified letter to all nonresponding units. The second studied adaptive design restricts the NRFU for the probability subsample of nonrespondents selected for the targeted allocation. In this paper, we compare the quality effects of the two studied adaptive NRFU designs examining effects on response, respondent sample balance, and collected data quality.

Key Words: Adaptive design, nonresponse subsampling, targeted allocation, embedded field experiment

1. Introduction

Currently, the Economic Directorate of the U.S. Census Bureau is conducting a series of embedded field experiments on data collection features in several ongoing annual business surveys. Strategies that prove to be successful can be quickly implemented into the annual programs, with an eye towards ultimately implementing them into larger periodic programs such as the Economic Census. Adaptive nonresponse follow-up (NRFU) strategies for small businesses are considered in this suite of improvement strategies. These adaptive strategies are motivated by "universal problems" in many ongoing programs. In recent years, survey researchers have faced declining response, increasing costs, and tighter budgets. Under these constraints, new, more tailored approaches to data collection that maintain the balance of quality and response, known as "adaptive" or "responsive" designs, have emerged. Groves and Heeringa (2006) first introduced the concept of responsive designs, a multi-phase approach to survey design that uses the outcomes of early data

Any views expressed on statistical issues or operational procedures are those of the authors and not necessarily those of the U.S. Census Bureau

collection to inform data collection strategies later in the field period. With a responsive design, researchers identify potential risks to quality and cost and make use of extant data, including sample frame data, paradata, response rates, and response data, to monitor data-collection and modify the contact strategy protocols throughout collection. In the simplest form, a responsive approach uses a two-phase design, identical to the two-phase sampling of nonrespondents proposed in Hansen and Hurwitz (1946) in which all sampled units might receive contact in an initial phase of data collection, but only a probability subsample of them receive contact in a second phase.

Researchers have many tools at their disposal for applying adaptive data collection strategies, which often include the use of experiments to implement and assess the changing strategies. These experimental applications include efforts to increase response rates through case prioritizations that manipulate the timing and frequency of contact with sampled households to increase response from low-propensity responders (Couper and Wagner, 2011; Peytchev et al, 2010; Wagner et al., 2012; Wagner, 2013). And although unit response rates are a common measurement outcome, researchers also have implemented adaptive strategies to address item-level quality by targeting both item nonresponse (Calinescu and Schouten, 2015) and final survey estimates (Beaumont, Bocci, and Haziza, 2014) for key survey variables.

However, because adaptive strategies seek to improve response, and implicitly reduce nonresponse bias, they often use collection strategies that are effective but impose higher costs and require higher levels of effort, such as using in-person interviewers for hard-to-reach cases instead of phone contact (Rosen et al., 2014). Yet, there is little evidence to suggest that an increase in response rates will always result in a reduction in nonresponse bias (Groves and Peytcheva, 2008). With the introduction of new methods for measuring potential nonresponse bias during data collection, such as R-indicators (Schouten, Cobben, and Bethlehem, 2009) or the balance and distance indicators (Särndal, 2011 and Särndal and Lundquist, 2014), researchers also have sought to balance response rates, implementation cost, and "representativeness" of the respondent sample. With R indicators, representativeness characterizes non-response that deviates from missing completely at random. A "balanced sample" is one where the actual respondents have the same or almost the same characteristics as the whole population for key measures (Särndal 2011). Schouten and colleagues have developed modeling schemas that seek to optimize the benefits of an adaptive strategy that increases both the response rates and representativeness within certain cost parameters (Calinescu, Bhulai, and Schouten, 2013; Luiten and Schouten, 2013; Schouten, Calinescu and Luiten, 2013, Särndal and Lundquist, 2014).

However, despite a growing body of literature on the use and success of adaptive strategies at improving quality and more effectively managing costs, much of this extant literature is focused on household surveys (see Tourangeau, et al. 2016 for a recent review).There is a paucity of published research on the use of responsive and adaptive designs for business surveys (see Wilson, McCarthy, & Dau, 2016 for one of the few examples). This lack of literature on adaptive designs in business surveys is not indicative of the fact that these approaches are not advantageous or appropriate for business surveys. Snijkers and colleagues (2013) advocate that business survey data collection should be tailored using optimal strategies to reach businesses and encourage participation. Instead, business survey data collection strategies are already inherently adaptive. The majority of collection strategies are aimed at obtaining response and quality data from the larger businesses, attempting to control both measurement error and nonresponse errors (see Willimack and Nichols 2010, Snijkers et al. 2013, Thompson and Oliver 2012, and Thompson and Washington 2013, among others).

The emphasis on larger businesses occurs because business populations are highly skewed, with a small proportion of sample units contributing to the majority of the industry totals. Because of the influence of these larger businesses on the estimates, business-survey data collection procedures are designed to increase the likelihood of obtaining responses from these larger businesses. For example, at the U.S. Census Bureau, the largest businesses, especially those that are surveyed in many different programs, are assigned Account Managers who maintain ongoing personal contact with the business (Brady 2016). These businesses are more likely to be contacted personally if there are questions about the data or as planned unit NRFU. In contrast, smaller businesses receive very little personal contact (if any). In general, smaller businesses are mailed reminder letters, but rarely receive telephone reminders or other personal contact. Accordingly, the unit response rates and total quantity response rates (item-level) for the large businesses included with certainty (sampled with probably = 1) are often well above the 70% benchmark recommended by the 2006 Federal Register Notice (Knutson and Cepluch 2016, Lineback and Fink 2012, Thompson and Oliver 2012, Thompson et. al 2015), whereas the same measures for the sampled (noncertainty) units tend to be below this benchmark. In all of the cited studies, the nonresponse from the small business components had a detrimental effect on the overall response rates. As response rates decrease across the board in many official statistics programs, this small business subpopulation becomes more important to the survey estimates, necessitating improvements in collection protocols.

Our experimental setting applies to programs that sample establishments (business locations), not companies (firms). Establishments fall into two broad categories: Single unit establishments own or operate a business at a single location and are classified into a single industry; and multi unit establishments are comprised of two or more establishments that are owned or operated by the same company. Companies in the second category receive one questionnaire per establishment and have to complete a variety of different forms depending upon the industries in which its establishments operate. For practical purposes, multi unit establishments are excluded from subsampling consideration to avoid compromising the extensive set of completeness procedures and to allow reconciliation with company level data for the same businesses collected in other surveys.

Single unit and multi unit establishments within the same industry can be quite similar in terms of size as measured by total sales, payroll, or employment, but the response burden and collection challenges are quite different. For example, Willimack and Nichols (2014) note that small businesses may not keep track of all the requested detailed data items. Bavdaž (2010) and Hedlin et al. (2008) note that small businesses may find the burden of responding to a survey as being too high. In contrast, multi unit companies are more likely

Any views expressed on statistical issues or operational procedures are those of the authors and not necessarily those of the U.S. Census Bureau

to maintain the detailed data on their records due to external regulations, but collecting the disaggregated establishment level data may be difficult (Willimack and Nichols 2014).

In this paper, we present the results of an embedded field experiment in the 2015 Annual Survey of Manufactures (ASM) that is the culmination of a four-year long exploration into applying adaptive collection strategies for establishment surveys. Our focus has been on one aspect of collection, specifically NRFU. Originally, we explored optimized allocation methodologies, under the erroneous assumption that clever sampling designs and carefully chosen estimation procedures could improve data quality without overly increasing sampling errors (Kaputa et al, 2014). This strategy proved to be misguided, providing evidence that the decrease in nonresponse bias was not offset by the increased sampling errors *unless* the nonrespondent subsampling was combined with an improved collection strategy. Unfortunately, there was limited research on effective contact strategies for small businesses to draw upon in the literature: exceptions include Hedlin et al. (2008) and Torres van Grinsven et al. (2014), although their target populations are somewhat different. That said, budget constraints often come into play in developing contact strategies for small businesses, as the majority of the contact strategy budget is allocated to the larger business.

In our quest for an effective NRFU collection strategy designed for small businesses, we conducted a field test to explore alternative protocols in the 2014 Annual Survey of Manufactures (ASM). The protocols included the certified mailing of a reminder letter, already proven to increase response rates for small business (Marquette et al. 2015), and an additional flyer written in a harsher tone recommended by the subject matter experts. Thompson and Kaputa (2017) discuss the findings from the earlier embedded experiment. That study reinforced the earlier findings in terms of improved response rates and length of collection time as well as data quality benefits for a few variables. Subsequently, certified reminder letter mailings for single units were implemented in the 2015 ASM collection and have been budgeted for single unit business NRFU in the upcoming 2017 Economic Census.

We selected the ASM as a testing ground for these adaptive NRFU procedures for several reasons. The ASM data collection strategy for single unit establishments is very similar to the Economic Census procedures. Both programs are mandatory, and sample units are informed of this at first contact. Both programs collect data from establishments. Furthermore, the ASM questionnaire, a subset of the manufacturing sector’s Economic Census questionnaire, is conducted in non-census years and uses the same editing and imputation procedures as the Economic Census. Thus, this controlled experiment allowed us to both examine quality effects of an adaptive design protocol in a controlled setting with strong similarities to the Economic Census and explore the feasibility and logistics of the protocol in a live survey. Even better, we had established procedures for implementing split panel tests with different NRFU collection strategies in the ASM from the 2014 experiment. Ideally, we would want to test Economic Census contact strategies in all economic sectors. Unfortunately, the other annual economic surveys conducted at the U.S. Census Bureau have different sample units (company versus establishment) and collect different items, making the extrapolation to the census a bit less transparent. And, not all

Any views expressed on statistical issues or operational procedures are those of the authors and not necessarily those of the U.S. Census Bureau

survey sponsors were comfortable with the risks associated with embedded experiments in stable ongoing programs.

The current experiment tested an adaptive NRFU collection strategy against the current fixed-design NRFU collection strategy. With the proposed adaptive design, a targeted probability subsample of nonrespondents receive a more expensive and effective NRFU collection strategy (certified reminder letter) with the unsampled units receiving an inexpensive protocol (regular mail reminder letter); with the current NRFU procedures, all nonresponding units receive a certified letter. The treatment panel is also used to simulate the outcome of pursuing only the nonresponding units selected in a probability subsample, similar to the method of Hansen and Hurwitz (1946). Using a split panel design allows us to examine different aspects of response including unit and item response rates, and potential nonresponse bias on selected items measured by the fraction of missing information (FMI) for key items (Thompson and Kaputa, 2017; Andridge and Thompson, 2015 (A and B), Wagner 2010). The evaluation statistics are presented in Section 2. Section 3 describes the study design and outlines the different NRFU collection strategies. It also presents the experimental design for the 2015 ASM field test. Section 4 presents the results. We conclude with a few specific observations about these studies along with general observations on the utility of embedded experiments in this and other similar settings.

2. Evaluation Statistics

In comparing adaptive NRFU procedures, we were interested in its effects on survey response and on the quality of the collected data. The first concern is primarily an administrative consideration. It is commonly accepted that unit response rates are often poor indicators of survey quality (Peytcheva and Groves 2009). However, federal programs are subject to the response rate guidelines presented in the Office of Management and Budget Statistical Standards These guidelines recommend that programs conduct nonresponse bias analyses studies when the unit response rate falls below 70%. Consequently, many programs, including the ASM, use the unit response rate as a performance benchmark.

The second question is more substantive. It is also far more difficult to address. The term "quality" has numerous definitions. Statistical agencies often define data quality as a combination of various measures, each examining different aspects. For example, Eurostat (2003) outlines five dimensions of data quality: relevance; accuracy and reliability; timeliness and punctuality; comparability and coherence; and accessibility and clarity (see http://ec.europa.eu/eurostat/web/income-and-living-conditions/quality). In this study, only the NRFU collection strategies change by panel, not the collection instrument. Consequently, the study results cannot be used to compare NRFU collection strategy effects on relevance, accessibility and clarity, or comparability and coherence. We could examine NRFU collection strategy effects on timeliness and did so in the earlier study (Thompson and Kaputa 2017), where we found some improvements using a certified letter mailing for the smallest businesses. More important, we can examine NRFU collection strategy effects on accuracy, with a few caveats. In the survey research literature, accuracy

Any views expressed on statistical issues or operational procedures are those of the authors and not necessarily those of the U.S. Census Bureau

is often defined in terms of total survey error (i.e. sampling error + nonsampling error). This is difficult if not impossible to measure. With this embedded experiment, this definition tends to predispose against subsampling nonrespondents, which increases sampling error by definition.

Instead, following Thompson, Oliver, and Beck (2015), we define "accuracy" as reporting accuracy in the sense the reported data values are not changed during the clerical and automatic review procedures. With this definition, we evaluate the accuracy of selected items under the different NRFU collection strategies using the quantity response rates (QRR) and source of data item (SDI) measures described below. We further define accuracy in terms of low nonresponse bias potential for selected variables, quantified by the fraction of missing information (FMI), as proposed in Wagner (2010) and implemented in Thompson and Kaputa (2017) for the earlier ASM field test. All evaluation statistics are computed by NRFU collection strategy (control/current procedure, targeted allocation, and nonrespondent subsampling) for all units and by the single unit subdomain. The NRFU strategies are applied to the single unit establishments in the split panel design, although the multi unit establishments are considered in the allocation procedures. Computing these metrics for all units provides indications of overall effects on the survey estimates of the alternative NRFU procedures; limiting measures to single unit establishments allows consideration of NRFU collection strategy effects without confounding but may tend to overstate such effects.

For business surveys, unit response rates (URR) are computed as unweighted ratios of respondents to eligible cases. This avoids overrepresentation of the smaller cases with larger weights in the response rate. In the official rate computations, a respondent is defined as an eligible reporting unit for which: (1) an attempt was made to collect data; (2) the unit belongs to the target population; (3) and the unit provided sufficient data to be classified as a response (Thompson and Oliver 2012). Unfortunately, for the current experiment, the collected ASM data have undergone only preliminary quality checks (edits) at the time of analysis, so we were unable to implement the third criterion. We use a check-in rate instead, satisfying only the first two criteria. For simplicity, we refer to the check-in rate as the "response rate."

Ultimately, survey stakeholders are concerned about NRFU collection strategy effects on response rate, particularly if those effects may be negative. By definition, response rates computed with subsampled respondents should be lower than those obtained from full follow-up. Of course, we hoped to see no NRFU collection strategy effects on final response rate between the control and targeted allocation panels. We use chi-squared tests for independence to test this hypothesis, accounting for the complex survey design using the Rao-Scott adjusted test in PROC SURVEYFREQ (SAS/STAT(R) 9.3 User's Guide 2015). The SAS procedure incorporates sampling weights, so that the tested response rates are different from the official unweighted measures. However, testing for differences without incorporating the complex survey design can lead to erroneous conclusions (Rao and Scott 1987). Moreover, the differences in corresponding weighted and unweighted response rates were trivial in our data sets.

Since smaller businesses tend to have large sampling weights, computing an unweighted rate reduces the influence of small businesses on the unit response rates but ignores the varying effects on the estimates of totals from the larger businesses. The Quantity Response Rate (QRR) addresses this deficiency. The QRR is the weighted proportion of an estimated total obtained from reported data. Unlike the unit response rate each data item has its own QRR, so there may be several QRR measures per survey. The results presented in Section 4 use the ASM design weight and include imputed values in the denominator estimates.

The QRR for an item approaches 100% when the majority of the estimated total is obtained from reported data. That said, if the largest businesses in the survey provide the majority of the tabulation data for an item, then the QRR could approach 100% while the realized unit response rate could be quite low if a high percentage of the smaller businesses did not respond. If there is no NRFU collection strategy effect, then we would expect that the QRR's for a given item would be approximately the same in each panel. We also examine accuracy effects from NRFU collection strategies by computing the Source of Data Item (SDI) statistic. The SDI measures the proportion of responding units that retain their reported data after processing (i.e. reported value equals edited value) for an item. Similar to the unit response rate, this proportion uses unweighted counts. We consider the SDI measures to be descriptive and do not conduct any formal testing for differences.

The QRR and SDI measures are useful quality metrics but provide limited insight into the potential for nonresponse bias. Andridge and Little (2011) observe that there are three components that can be used to assess the potential for nonresponse bias: the amount of nonresponse, the differences between respondents and nonrespondents on fully observed characteristics (e.g., paradata, frame data), and the relationship between these fully observed characteristics and the survey outcomes (only measureable among respondents). The URR analyses examine the first component, and the QRR and SDI examine the third component.

The challenge lies in the second component. We could systematically examine differences between respondents and nonrespondents using frame data variables. For example, the distance measure proposed by Särndal and Lundquist (2014) compares the difference between mean value for respondents and mean value for nonrespondents on variable x, available for all units. With our datasets, the distance measures proved unenlightening when computed with payroll as all measures were essentially equal to the nominal value of zero.

Instead, we developed proxy pattern-mixture (PPM) models, following the procedure recommended in Andridge and Little (2011) and the methods developed for skewed data distributions described in Andridge and Thompson (2015A and B). The FMI has been proposed as a metric for assessing the risk of nonresponse bias for a specific adjusted survey estimate (Wagner 2010, Wagner 2012, Andridge and Little 2011, Andridge and Thompson 2015 (A)), measuring the loss of precision due to nonresponse after imputation (adjustment).To conduct these analyses, we create a single "proxy" variable X for every studied outcome variable by regressing the outcome variable on the items used for imputation, nested within imputation cell (a no-intercept linear regression model). The joint distribution of a survey outcome Y and this proxy X is modeled as a bivariate gamma

Any views expressed on statistical issues or operational procedures are those of the authors and not necessarily those of the U.S. Census Bureau

distribution (appropriate for a skewed manufacturing population) with separate parameters for respondents and nonrespondents (a pattern-mixture model).

In the PPM framework, the FMI values are strongly related to the strength of the predictors used in the proxy. In some instances, a NRFU collection strategy effect might be completely ameliorated by an excellent predictor (strong proxy). The converse can also be true if the relationship between predictors and outcomes is not strong (weak proxy). These PPM FMIs are computed with respect to a specified imputation model, so that an FMI value close to zero indicates little or no nonresponse bias effects in the variable after adjustment and a value close to one indicates the reverse.

To test the robustness of the imputation model to response mechanism, we computed the FMI of a given outcome variable at the two extremes, specifically missing at random (MAR) and not missing at random (NMAR). If the FMI values for the variable obtained under different response mechanisms are close together, then the inflation of variance due to an NMAR mechanism is not severe, relative to the MAR mechanism. For a more detailed discussion of the factors impacting FMI and its use in the PPM framework, see Andridge and Thompson (2015A).

Here, we are particularly interested in assessing whether one (or both) of the NRFU collection strategies has a detrimental impact on the FMI of one or more variables and if there are small differences within treatment panel in corresponding FMI estimates (MAR vs. NMAR). If the targeted allocation approach or nonrespondent subsampling are yielding comparable realized response sets to the current procedure, then all three FMI values (one per treatment panel) should be approximately the same for each outcome variable.

3. Field Test Design

3.1. Annual Survey of Manufactures (ASM) Survey Background

The purpose of the ASM is to produce “sample estimates of statistics for all manufacturing establishments with one or more paid employees.” The ASM collects general manufacturing statistics including total payroll, number of employees, receipts (shipments), and total hours worked by production workers. The survey is conducted annually in years between the Economic Census, with a new fixed Pareto-PPS sample selected two years after the most recent Economic Census; the Economic Census is conducted in years ending with a 2 and 7. Approximately 50,000 establishments are selected from a universe of nearly 297,000 establishments; of those, 15,600 establishments are included with certainty (sampled with probability =1) and the remaining establishments are selected with probability proportional to a composite measure of size based on Economic Census shipments. The ASM imputes a complete record for unit nonrespondents. For more details on the ASM design and estimation procedures, see the ASM website at https://www.census.gov/programs-surveys/asm.html.

Typical of many business surveys, the ASM phone follow-up procedures focus on obtaining respondent data from the largest businesses. The multi units and largest single units are consequently a higher priority for follow-up. All the remaining nonresponding cases receive mail reminders, resulting in the small single unit establishments the least likely to receive personal contact as follow-up. At the time of this experiment, all sample units received an initial contact letter requesting internet response; this is then followed by a due date reminder. The first round of NRFU is a reminder letter, a second round follows with a certified reminder letter, and finally an extremely strongly worded reminder letter emphasizing the mandatory nature of the program at the third round.

Nonrespondent subsampling (and by extension, targeted NRFU allocation) was restricted to the ASM single unit nonrespondents, although all sampled units (multi and single) are included in the domain response rate estimates used for allocation. This parallels the protocol under consideration in the 2017 Economic Census. Restricting the eligible-unit domain to single unit establishments sidesteps any modifications to completeness and coverage procedures with the multi unit establishment universe. More important for the ASM, it greatly reduces potential impact on reliability restrictions. In fact, there are no sampling error increases when the targeted allocation NRFU procedure is implemented.

3.2. Experimental Design

To ensure sufficient sample in the experiment, there were two experimental panels: a panel implementing the targeted NRFU design and a control panel where every nonresponding unit receives the same follow-up procedures. The nonrespondent subsampling is "simulated" from the targeted allocation by flagging the non-targeted units as nonrespondents and only including the targeted/subsampled units' responses as valid. This allowed us to simultaneously compare two different treatments and maximize the size of the treatment panel. We are aware that some of the non-subsampled units could potentially respond before survey close-up, but we assume the worst case scenario of zero respondents for this study. With this experiment, all ASM single unit establishments received the same initial contact letter, due date reminder letter, and 1st NRFU reminder letter. This maximizes the usage of previously proven contact strategies. Before the 2nd NRFU, we paired three digit industries based on a standardized nearest neighbor distance available from the SAS MODECLUS procedure (SAS/STAT(R) 9.3 User's Guide 2015) calculated on industry sample size, current estimated response rate, total frame payroll, and total weighted frame payroll. One industry in each pair was randomly assigned to each panel. All nonresponding eligible (single unit) establishments in the control panel receive the standard follow-up (certified letter and additional reminder letter if necessary). Figure 1 provides a diagram of the experimental design.

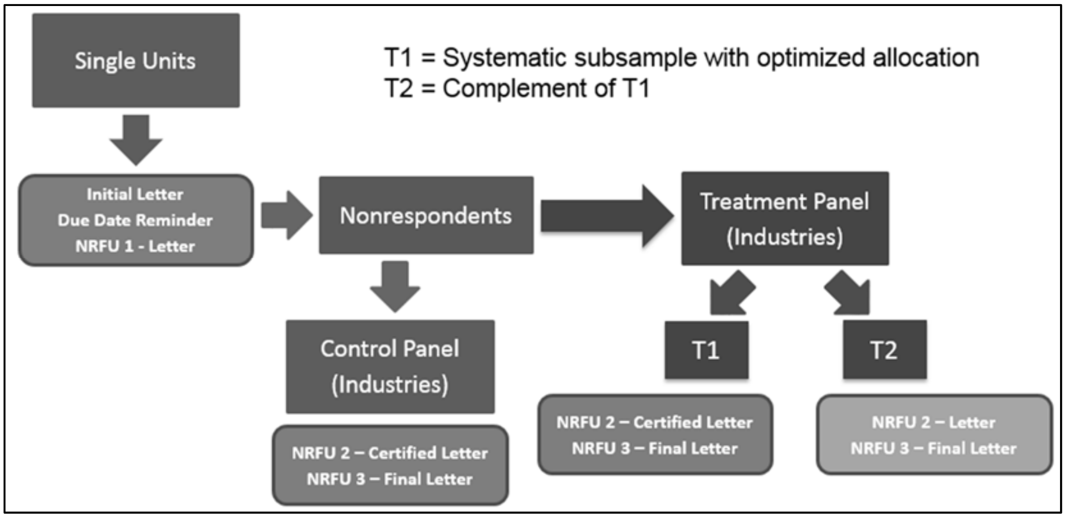

Figure 1: Diagram of the Experimental Design for the 2015 ASM Field Test

Figure 2 illustrates the NRFU collection strategies for the experiment. The first block represents the control panels. Establishments that provided a valid response before the 2nd NRFU procedure are denoted as respondents. The number of respondents varies by domain (industry in our application). All nonresponding units received a certified letter and an additional reminder letter if necessary. This is the most expensive NRFU collection strategy considered in the field test ("$$"). The second block illustrates the targeted allocation treatment. For this, we selected a systematic sample of single unit establishments from the frame of nonresponding units in the treatment panel industries, determining the sample size of the more expensive (certified letter) protocol via the optimized allocation method described in the Appendix. These subsampled/targeted units received a certified letter and an additional reminder letter if necessary. The "unsampled" nonrespondents in the treatment panel received a noncertified (regular mail) letter and an additional reminder letter if necessary. The latter procedure is less expensive ("$"). Because the nonrespondent subsampling allocation procedure is designed to select larger subsampling in low-responding domains, the proportions of "$$" and "$" differ by domain. Lastly, the third block illustrates nonrespondent subsampling.

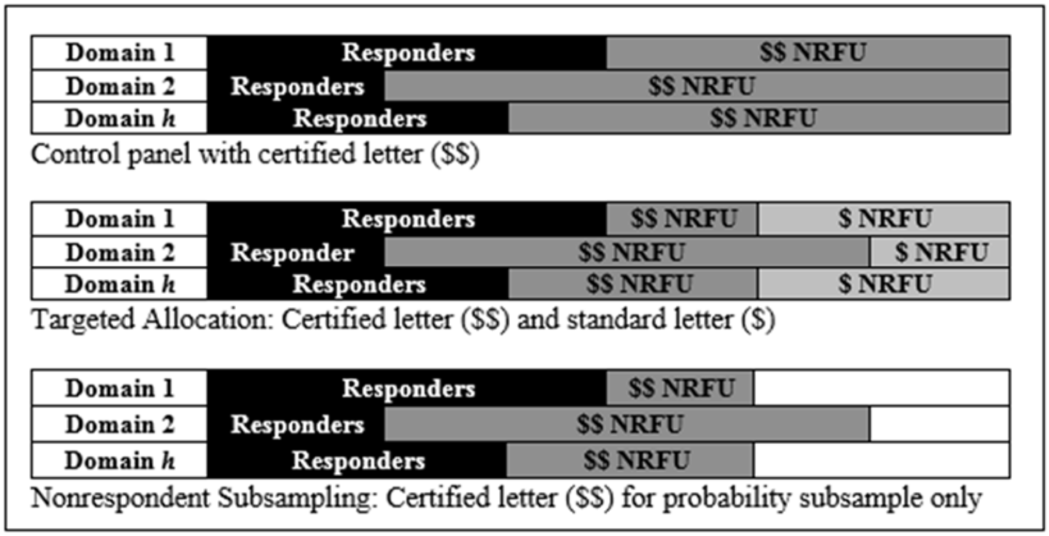

Figure 2: NRFU Collection Strategy Comparisons for the 2015 ASM Field Test

For budgetary and reliability reasons, we selected an overall 1-in-2 subsample of nonrespondents from the treatment industries. The NRFU costs from the targeted allocation are less than the control panel, as only half of the nonrespondents receive certified letters. Nonrespondent subsampling is less expensive than targeted allocation, as there are costs associated with the regular mail reminders sent to the unsampled nonrespondents units.

Table 1 contains the estimated parameters for the allocation and the resulting subsampling rates; see the Appendix for the optimal allocation algorithm. The table also contains the final predicted response rates for the given allocation. Establishment response status at allocation was provided by ASM subject matter experts.

The historical conversion rates (the conditional probability of responding at a given point in time) used for the allocation were estimated from the 2014 ASM data. Taking these conversion rates into account along with the maximum allowable subsample (1-in-2), the target response rate for all domains was 69 percent. Table 1 provides a complete picture of the subsampling. The first set of columns ("Response Rate Prior to Nonrespondent Subsampling") provide the response rates at the time of subsampling, classified by eligibility status. The second set of columns provide the historic conversion rates (q_h) at the time of subsampling (conditional on not responding previously), again classified by eligibility status(q_h^i *and* q_h^e). The allocation response rates provided in Table 1 below are the starting for the optimal allocation. These allocation response rates are calculated as

$$URR_h^A = \frac{r_{1h} + (m_{1h}^i * q_h^i)}{n_h}$$

where r_{1h} are *all* responding establishments in domain h immediately prior to subsampling (eligible and ineligible units) and m_{1h}^i are the *ineligible* nonrespondents in the domain (for subsampling). A subsampling rate of 1 is full NRFU (100% follow-up) with all units receiving the most expensive NRFU collection strategy; a subsampling rate of 0 implies no subsampling (if nonrespondent subsampling is implemented) or that all eligible units in the domain receive the inexpensive protocol.

Table 1: Nonrespondent subsample allocations with a unit response rate target of 69 percent for the 2015 ASM

Industry	Response Rate Prior to Nonrespondent Subsampling			Historical Conversion Rate		Allocation Response Rate	Subsample Rate	Predicted Final Response Rate for Nonrespondent Subsampling		
	All	Ineligible	Eligible	Ineligible	Eligible			All	Ineligible	Eligible
1	31%	10%	51%	0.61	0.48	58%	0.94	69%	65%	73%
2	23%	9%	54%	0.61	0.58	61%	0.94	69%	64%	79%
3	34%	10%	58%	0.61	0.48	61%	0.77	69%	65%	73%
4	41%	9%	62%	0.59	0.57	62%	0.53	69%	62%	73%
5	18%	10%	56%	0.62	0.51	63%	1.00	68%	65%	78%
6	38%	9%	63%	0.65	0.58	65%	0.33	69%	68%	70%
7	19%	10%	57%	0.66	0.54	67%	0.48	70%	70%	68%
8	26%	19%	54%	0.65	0.53	68%	0.46	70%	72%	65%
9	26%	11%	59%	0.74	0.54	71%	0.00	71%	77%	59%
10	21%	21%	35%	0.82	0.44	84%	0.00	84%	86%	35%

Notice how industries 9 and 10 both start with allocation response rates above the target response rate, eligible units will receive a less expensive follow-up in these domains. On the opposite end of the spectrum, all of the eligible nonresponding units in industry 5 receive the more expensive NRFU collection strategy because their predicted final response rate falls under the target (given full NRFU). The remaining industries have subsampling rates ranging from as low as 1-in-3 to as high as 1-in-1.064.

4. Results

Figure 3 presents unit response rates for all ASM by NRFU collection strategies over time. Figure 4 presents the corresponding measures computed for single unit cases (experimental population). The experimental treatments begin with NRFU 2. Regardless of panel, all units that have not provided a response by NRFU 3 receive the same final reminder letter.

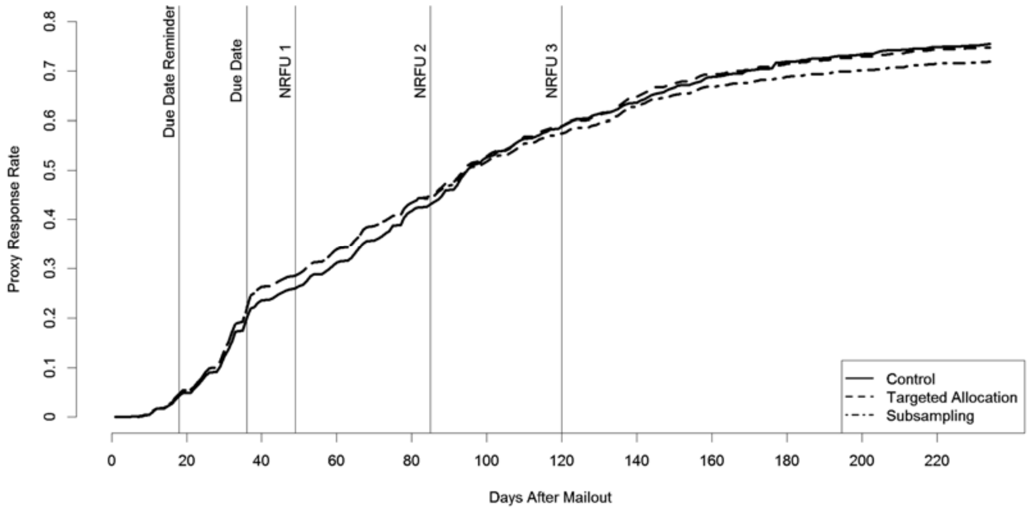

Figure 3: Proxy Response Rates (Check-in rates) for All Cases (Multi Unit and Single Unit Establishments) Over Time for the 2015 ASM

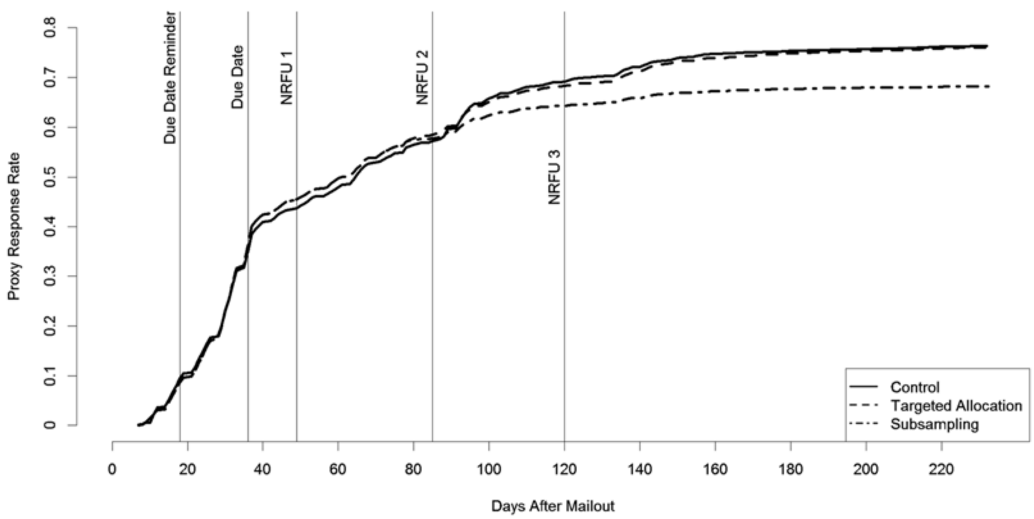

Figure 4: Proxy Response Rates (Check-in rates) for Single Unit Cases (Experimental Population) Over Time for the 2015 ASM

In both graphs, the control panel and targeted allocation panel response rates are indistinguishable: for all units, the final response rates are 75.72% (control) and 75.09% (targeted allocation) and are not significantly different (p-value = 0.65); for the single unit cases, the final response rates are 76.32% (control) and 76.08% (targeted allocation) and are likewise not significantly different (p-value = 0.76). This provides evidence that the less expensive targeted allocation approach is as effective in eliciting response as the more expensive currently implemented procedure (control panel) due to the selective mailing of certified letters. Notice that the response rates for the subsampled treatment group are significantly lower than the other two, even when multi unit establishments are included in the computations.

Table 2 provides the QRR by collection strategy for four key outcome variables: payroll, employment, receipts, and production worker hours. Of the collected variables, payroll is the most heavily scrutinized, as the remaining key items are edited with respect to the (validated) payroll value for each of these variables by NRFU collection strategy. Values that are significantly different from the control panel results at $\alpha = 0.05$ are indicated by an asterisk.

Table 2: Quantity Response Rates by NRFU collection strategy for the 2015 ASM

Item	Domain	Control Panel	Targeted Allocation	Subsampled Units
Payroll	All	67.34%	68.56%*	68.55%*
	SU	84.90%	84.27%	83.73%
Employment	All	78.76%	79.31%	79.23%
	SU	79.39%	77.80%	77.55%
Receipts	All	83.85%	81.21%*	81.19%*
	SU	87.50%	91.80%*	91.96%*
Product Worker Hours	All	51.53%	83.19%*	81.58%*
	SU	76.25%	84.48%*	80.01%*

With payroll and employment, there is no evidence of a difference in QRR in the single unit (SU) domains that actually received different NRFU collection strategies. With receipts and production worker hours, the targeted allocation QRR values are larger than their control panel counterparts in the single unit domains. We caution against drawing general conclusions from this result, as it could be an artifact of the industry pairing in the experimental design if one of the industries has notably poorer response for the studied items.

Table 3 presents the SDI by NRFU collection strategy for the four outcome variables. Due to the extensive pre-editing of payroll, the apparently low reported data retention rates for this variable of approximately 85% are not unexpected. Moreover, the two NRFU collection strategies (control/uniform treatment and targeted allocation) yield very similar rates.

Table 3: Source of Data Item by NRFU collection strategy for 2015 ASM

Item	Domain	Control Panel	Targeted Allocation	Subsampled Units
Payroll	All	82.36%	84.17%	84.22%
	SU	83.74%	84.39%	84.52%
Employment	All	80.65%	78.41%	78.48%
	SU	76.41%	74.33%	74.05%
Receipts	All	88.32%	88.60%	88.56%
	SU	88.43%	90.48%	90.55%
Product Worker Hours	All	81.29%	78.06%	78.30%
	SU	74.88%	71.87%	71.84%

The SDI provides indications of the data quality of the collected items, given a NRFU collection strategy. Subtle differences in SDI could be attributed to random error, a consequence of the paired design, or a systematic effect (e.g., cases that receive a certified reminder are more likely to complete the entire questionnaire because the treatment increases their perception of their value in the survey). However, all establishments contribute equally to the SDI regardless of unit size. If larger businesses are providing the high quality reported data, then the QRR measures should be close to 100% as well.

Item level rates, QRRs and SDIs were comparable for both experimental procedures. Moreover, the program managers were uncomfortable with the strict nonrespondent subsampling procedure and endorsed the targeted allocation procedures. Consequently, we dropped the subsampling protocol from further consideration.

To study the quality effects of the two alternative treatments on specific items, we conducted proxy pattern-mixture (PPM) analysis, comparing the FMI between the control group and the targeted allocation for four collected items (payroll, employees, receipts, and hours worked). We model a separate proxy for each collected item and treatment panel by regressing the survey variable on the frame variable payroll within 3-digit industry (no intercept), using multiple imputation with 200 draws given a burn-in period of 500 draw and thinning at every 10th draw. The models used to develop payroll and employment proxies use all the respondent data; outliers were removed from the receipt and hours worked data to improve the fit of the regression model used for the proxy.

Figure 4 presents the adjusted-R^2, unit nonresponse rates, and FMI for each treatment panel by outcome variable. The FMI is presented as range, with the lower limit representing the value obtained assuming a MAR response mechanism (the best case scenario) and the upper value representing the value obtained assuming a NMAR response mechanism (the worst case scenario). The FMIs for the complete survey (all establishments) are presented on the left; the FMIs for the experimental population (single units only) are presented on the right.

All Units Experimental Population (Single Units)

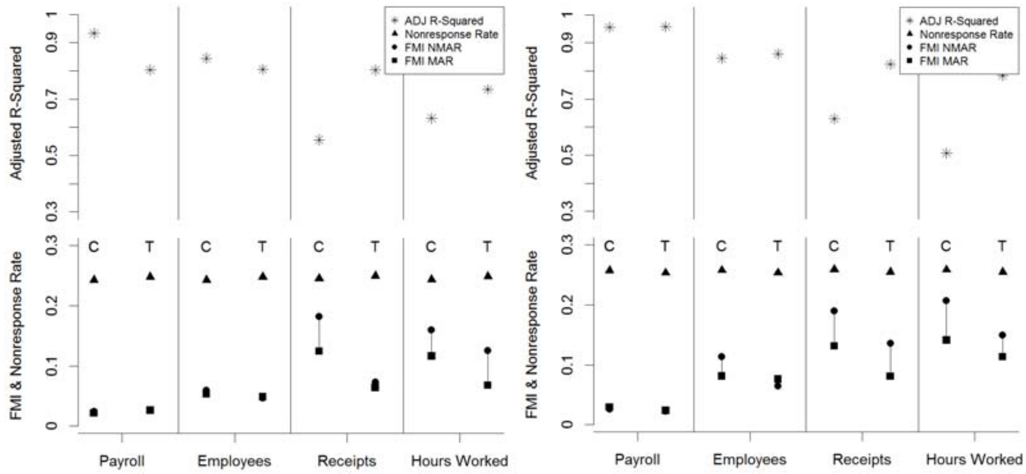

Figure 4: FMI for all units and single units for four key survey variables by treatment panel under two alternative response mechanisms for the 2015 ASM (C = Control Group, T = Targeted Allocation).

The adjusted-R^2 values provide information on the predictive strength of the proxy. The frame variable payroll is expected to be strongly correlated with the studied survey outcome variables. The correlation between collected payroll and frame payroll is near unity since frame payroll is a estimated from prior year ASM payroll. Within industry, employment is usually strong predictor of payroll, and the proxy model for total employees is very strong as well, albeit slightly weaker than the payroll proxy. Note that the proxies for both receipts and hours worked are much weaker than their payroll and employment counterparts. Payroll, employment, and receipts are generally well-reported variables. Moreover, they are highly inter-correlated, so that it is not difficult to develop strong imputation (prediction) models for any one of these items within industry. In contrast, production work hours tends to poorly reported, especially among single unit establishments (the experimental population).

For payroll and employment, the strength of the proxies are comparable between corresponding experimental panels. However, they are not comparable for receipts nor for hours worked. The adjusted-R^2 values are notably higher in the treatment panel for these variables, which reinforces our earlier stated concerns about the differences in QRR and SDI for these values.

Andridge and Little (2011) recommend comparing the range of the FMI values to the nonresponse rates to examine the “severity of nonresponse for a particular outcome.” In this experimental setting, we expect that the FMI for an item should *decrease* as the “representativeness” or “balance” of the realized sample increases, using the cited definitions in Section 1. In our setting, having all the FMI values fall below the nonresponse rate provides further evidence that the realized set of respondents are a random subsample of the original sample and preliminary evidence that the survey estimate for these items would not be overly influenced by nonresponse bias. With that said, all FMI value ranges are well below the nonresponse rate, regardless of population (all units or the experimental population), experimental treatment, or item.

Because the payroll proxy is so strong in the experimental population of single unit establishments, the FMI values are approximately zero for this variable, regardless of assumed response mechanism. Intuitively, this makes sense. However, the targeted allocation NRFU collection strategy (treatment panel) yields the lowest FMI for three of the four studied items. Furthermore, the range of FMI under varying response mechanisms is smaller. With employees, the differences in FMI and FMI ranges between panels is less pronounced than for receipts and for hours worked, although it is evident. Unfortunately, the differences in FMI and FMI ranges for receipts and hours worked are confounded with the differing proxy strengths by panel.

The “cleanest” comparisons between FMI and FMI range can be made in the experimental population (single units only) with the payroll and employee items, as including the multi unit cases can mitigate treatment differences. Here, the results are indicative of improved respondent sets over the current procedure with resultant data quality improvements. However, more research is needed to support this conclusion.

5. Conclusion

The objective of this field test was to find a NRFU collection strategy that achieved comparable response and data quality, at a reduced cost, from a selected population of small businesses. The proposed adaptive collection strategy builds upon previous research on subsampling and collection strategies. Borrowing the most effective allocation strategy from a prior study (Kaputa et al 2014), we obtained a probability subsample of nonrespondents that would receive the most effective (and expensive) NRFU collection strategy determined by a second study (Thompson and Kaputa 2017). However, instead of implementing nonrespondent subsampling, all units in the treatment panel receive some form of NRFU.

Field experiments previously had proven valuable for developing a viable NRFU collection strategy for small businesses in the ASM. The 2014 ASM field test provided convincing evidence that uniformly mailed certified reminders increased response over regular mail and led to immediate changes in the ASM NRFU collection strategy for small businesses. The earlier test required strong support from the program managers to implement. Fortunately, the program managers were reassured by the final survey unit response rate and believed the experimental evidence. They were very supportive of the subsequent experiment presented here, especially since it had the added benefit of reducing overall survey costs in the treatment industries.

With this field test, we suspected that the targeted allocation strategy could likewise reduce nonresponse bias and consequently balance the respondent sample. Of course, nonrespondent subsampling is even less expensive than the targeted allocation procedure. Since we selected a probability subsample of nonrespondents for the targeted allocation, we believed that nonrespondent subsampling – like targeted allocation – would not have a detrimental effect on data quality. However, we were concerned that the cost savings would be offset by the decreased response rates and increased sampling variances.

The targeted allocation strategy worked, yielding similar response rates to uniformly mailed certified reminders at a reduced cost. More importantly, there was no evidence of degraded data quality in any of the studied variables and were even some improvements in data quality. While the latter could be coincidental, we caution against discounting the comparability. The optimal allocation and sample selection programs run quickly, and the adaptive NRFU procedure proved easy to implement. Given that, not only did we recommend continuing to adopt this strategy in future ASM collections, but also we suggested that it be implemented on a larger scale in other programs. The Economic Census stakeholders agreed, and implementation planning is well underway for the 2017 census.

In our experience, the "one-size-fits-all" NRFU collection strategy is overly heavy-handed. Although our recommended NRFU collection strategy may not be generalizable to other business surveys – indeed, the Economic Census may choose to modify the procedures – the strategic targeted allocation approach should be generalizable. Intuitively, it makes sense to concentrate expensive NRFU efforts in low-responding or perhaps even in potentially high bias domains. Likewise, it makes sense to *not entirely* discontinue NRFU in high responding domains; after all, survey designs usually have a target sample size and response rate. There are cost savings with our proposed approach, although the amount saved will depend on the implemented collection strategies. Most importantly, these results demonstrate that an adaptive collection design can be easily implemented without any deleterious effects on response rate or to data quality.

This sequential and piecemeal experimental approach to developing the final recommended adaptive collection design for small businesses has been educational and fruitful. Mirroring the practices in the biological sciences, we tested one new feature at a time in the same survey, avoiding confounding to the extent possible. Systematically dropping procedures or methods that did not demonstrate improvements in consecutive experiments while adopting the proven methods made it easier to identify causality. This sequential experimental approach is not atypical of household surveys, but can be lacking in establishment surveys where the tendency is to test several new elements simultaneously instead of using the more conservative factorial design.

As we noted, business survey collection strategies tend to be inherently adaptive. However, the focus is usually on the large businesses. With a probability sample, all units represent a component of the target population. Developing effective NRFU collection strategies for small businesses reduces the (unmeasurable) nonresponse bias in the survey estimates. Even being constrained to relatively inexpensive methods compared to personal telephone calls or field visits, we can achieve moderate cost savings and quality improvements for small businesses with alternative adaptive collection strategies.

References

Any views expressed on statistical issues or operational procedures are those of the authors and not necessarily those of the U.S. Census Bureau

Andridge, R.R. and R.J.A. Little. 2011. Proxy Pattern-Mixture Analysis for Survey Nonresponse. *Journal of Official Statistics* 27: 153-180. Available at: http://www.jos.nu/Articles/abstract.asp?article=272153.

Andridge, R.R. and K.J. Thompson. 2015(A). Using the Fraction of Missing Information to Identify Auxiliary Variables for Imputation Procedures via Proxy Pattern-Mixture Models. *International Statistical Review* 83(3): 472–492. DOI: 10.1111/insr.12091.

Andridge, R.R. and K.J. Thompson. 2015(B). Assessing Nonresponse Bias in a Business Survey: Proxy Pattern-Mixture Analysis for Skewed Data. *Annals of Applied Statistics* 9(4): 2237–2265. DOI: 10.1214/15-AOAS878.

Bavdaž, M. 2010. The Multidimensional Integral Business Survey Response Model. *Survey Methodology* 36: 81- 93.

Beaumont, J.F., Bocci, C, and Haziza, D. 2014. An Adaptive Data Collection Procedure for Call Prioritization. *Journal of Official Statistics* 30(4): 607-621.

Brady, C. 2016. "Respondent Outreach Practices at the U.S. Census Bureau." In Proceedings of the Fifth Conference on Establishment Surveys (ICES-V), June 6, 2016. Alexandria, VA: American Statistical Association.

Calinescu, M., S. Bhulai and B. Schouten. 2013. Optimal Resource Allocation in Survey Designs. *European Journal of Operations Research* 226(1): 115–121.

Calinescu, M. and Schouten, B. 2015. Adaptive Survey Designs to Minimize Survey Mode Effects –A Case Study on the Dutch Labor Force Survey. *Survey Methodology* 41 (2): *403-425*.

Couper, M. and J. Wagner. 2011. Using Paradata and Responsive Design to Manage Survey Nonresponse. *Proceedings of the World Statistics Congress of the International Statistical Institute.*

Federal Register Notice. 2006. OMB Standards and Guidelines for Statistical Surveys.

Groves, R. M. and S.G. Heeringa. 2006. Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs. *Journal of the Royal Statistical Society Series A* 169: 439–457. DOI:10.1111/j.1467-985X.2006.00423.x

Groves, R.M. and E. Peytcheva. 2008. The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. *Public Opinion Quarterly* 72 (2): 167-189. doi: 10.1093/poq/nfn011

Hansen, M.H. and Hurwitz, W.N. 1946. The Problem of Non-Response in Sample Surveys. *Journal of the American Statistical Association* 41: 517-529.

Hedlin, D., H. Lindkvist, H, Bäckström, and J. Erikson. 2008. An Experiment on Perceived Survey Response Burden Among Businesses. *Journal of Official Statistics* 24(2): 301-318.

Kaputa, S.J., L. Bechtel., K.J. Thompson, and D. Whitehead. 2014. Strategies for Subsampling Nonrespondents for Economic Programs." In *Proceedings of the Section on Survey Research Methods*, August 6, 2014. Alexandria, VA: American Statistical Association. Available at: http://ww2.amstat.org/sections/srms/Proceedings/ (accessed February 2017).

Knutson, J. and G. Cepluch. 2016. Nonresponse Bias Analysis for the U.S. Census Bureau's Quarterly Financial Report. In *Proceedings of the Section on Government Statistics*: American Statistical Association. Alexandria, VA: American Statistical Association.

Any views expressed on statistical issues or operational procedures are those of the authors and not necessarily those of the U.S. Census Bureau

Little, R.J.A. and D.B. Rubin. 2002. Statistical Analysis with Missing Data (2nd Edition). New York: Wiley.

Lineback, J.L. and E. Fink. 2012. Recent Developments in Assessing and Mitigating Nonresponse Bias. In *Proceedings of the Fourth International Conference on Establishment Surveys (ICES-IV).* June 13, 2012. Alexandria, VA: American Statistical Association. Available at: *www.amstat.org/meetings/ices/2012/papers/302146.pdf* (accessed February 2017).

Luiten, A. and B. Schouten. 2013. Adaptive Fieldwork Design to Increase Representative Household Survey Response; a Pilot Study In The Survey Of Consumer Satisfaction. *Journal of Royal Statistical Society, Series A* 176(1): 169–190.

Marquette, E., M. Kornbau, and J. Toribio. 2015. "Testing Contact Strategies to Improve Response in the 2012 Economic Census. In *Proceedings of the Section on Government Statistics*: American Statistical Association, August 10, 2015. Alexandria, VA: American Statistical Association.

Peytchev, A., E. Peytcheva, R.M. Groves. 2010. Measurement Error, Unit Nonresponse, and Self-Reports of Abortion Experiences. *Public Opinion Quarterly* 74 (2): 319-327. doi: 10.1093/poq/nfq002.

Peytcheva, E. and R.M. Groves. 2009. Using Variation in Response Rates of Demographic Subgroups as Evidence of Nonresponse Bias in Survey Estimates. *Journal of Official Statistics* 25:193-201.

Rao, J.N.K. and A.J. Scott. 1987. On Simple Adjustments to Chi-Square Tests with Sample Survey Data. *The Annals of Statistics* 15(1): 385-397.

Rosen, J., Murphy, J., Peytchev, A., Holder, T., Dever, J., Herget, D., & Pratt, D. 2014. Prioritizing Low-Propensity Sample Members in a Survey: Implications for Nonresponse Bias. *Survey Practice*, 7(1): 1-8.

"SAS/STAT(R) 9.3 User's Guide." *SAS/STAT(R) 9.3 User's Guide*. N.p., n.d. Web. 09 Oct. 2015.

Särndal, C.E. 2011. "The 2010 Morris Hansen Lecture: Dealing with Survey Nonresponse in Data Collection in Estimation." *Journal of Official Statistics* 27: 1-21.

Särndal, C. and P. Lundquist. 2014. Accuracy in Estimation with Nonresponse: A Function of Degree of Imbalance and Degree of Explanation. *Journal of Survey Statistics and Methodology* 2(4): 361-387.

Schouten, B., F. Cobben, and J. Bethlehem. 2009. Indicators for the Representativeness of Survey Response. *Survey Methodology* 35: 101-113.

Schouten, B., M. Calinescu, M., and A. Luiten. 2013. Optimizing Quality of Response through Adaptive Survey Designs. *Survey Methodology* 39(2): 29-58.

Snijkers, G., G. Haraldsen, J. Jones and D.K. Willimack. 2013. Designing and Conducting Business Surveys. Hoboken, NJ: John Wiley & Sons, Inc.

Thompson, K.J. and Kaputa, S. 2017 (accepted). Investigating Adaptive Nonresponse Follow-up Strategies for Small Businesses through Embedded Experiments. *Journal of Official Statistics.*

Thompson, K.J. and B.E. Oliver. 2012. Response Rates in Business Surveys: Going Beyond the Usual Performance Measure. *Journal of Official Statistics* 28: 221-237. Available at: http://www.jos.nu/Articles/abstract.asp?article=282221.

Thompson, K.J., B.E. Oliver and J. Beck. 2015. An Analysis of the Mixed Collection Modes for Two Business Surveys Conducted by the US Census Bureau. *Public Opinion Quarterly* 79 (3): 769-789. DOI: 10.1093/poq/nfv013

Thompson K. J. and K.T. Washington. 2013. Challenges in the Treatment of Unit Nonresponse for Selected Business Surveys: A Case Study. *Survey Methods: Insights from the Field.* Retrieved from http://surveyinsights.org/?p=2991.

Tourangeau. R., Brick, J.M., Lohr, S., and Li, J. 2016. Adaptive and Responsive Survey Designs: a Review and Assessment. *Journal of the Royal Statistical Society Series A* 180: 203-223.

Torres van Grinsven, V., Bolko, I., and Bavdaž, M. 2014. In Search of Motivation for the Business Survey Response Task. *Journal of Official Statistics*, 30(4), pp. 579–606.

Wagner, J. 2010. The Fraction of Missing Information as a Tool for Monitoring the Quality of Survey Data. *Public Opinion Quarterly* 74: 223–243. DOI: 10.1093/poq/nfq007.

Wagner, J. 2012. A Comparison of Alternative Indicators for the Risk of Nonresponse Bias. *Public Opinion Quarterly* 76 (3): 555-575. DOI: 10.1093/poq/nfs032.

Wagner, J. 2013. Adaptive Contact Strategies in Telephone and Face-to-Face Surveys. *Survey Research Methods* 7(1):45-55.

Wagner, J., N. Kirgis, B. West, J. Lepkowski, W. Axinn, and S. Kruger-Ndiaye. 2012. Use of Paradata in a Responsive Design Framework to Manage a Field Data Collection. *Journal of Official Statistics* 28(4):477-499.

Willimack, D. and E. Nichols. 2010. A Hybrid Response Process Model for Business Surveys. *Journal of Official Statistics* 26: 3-24.

Wilson, T., McCarthy, J., & Dau, A. 2016. Adaptive Design in an Establishment Survey: Targeting, Applying and Measuring 'Optimal' Data Collection Procedures in the Agricultural Resource Management Survey. In *Proceedings of the Fifth International Conference on Establishment Surveys (ICES-V)*. Geneva, Switzerland: American Statistical Association.

Any views expressed on statistical issues or operational procedures are those of the authors and not necessarily those of the U.S. Census Bureau

Appendix: Optimal Allocation Used in the ASM Experiment

Given a probability sample of n units divided into h disjointed domains, our goal is to select a probability subsampling of nonrespondents within domain for targeted allocation under the following conditions:

- The domains should have approximately the same sample sizes before unit nonresponse
- Each domain may contain units that will be excluded from subsampling (ineligible units)

Although the ineligible units cannot be subsampled, they are included in the domain response rate measures for the optimal allocation.

After all units receive some form of initial contact, each domain h will contain r_{1h} responders and m_{1h} nonresponders. We divide the m_{1h} nonresponders into m_{1h}^{i} ineligible nonresponders and m_{1h}^{e} eligible nonresponders. We obtain the subsampling rate for each domain (K_h) to select a systematic subsample of eligible nonrespondents in each domain h, resulting in m_{2h}^{e} selected units within domain for targeted follow-up.

We formulate the allocation as a quadratic program that minimizes the squared deviation in domain subsampling rates (K_h) to a fixed overall subsampling rate (K), with additional constraints on unit response rate and subsample size. The objective function is:

$$\min \sum_{h} \left(K_h \ - \mathrm{K}\right)^2$$

We add the notation below to define the additional constraints on unit response rate.

q_h^e = Conversion rate for eligible nonresponders in domain h
q_h^i = Conversion rate for ineligible nonresponders in domain h
m_h^e = Total count of eligible nonresponders in domain h
m_h^i = Total count of ineligible nonresponders in domain h

The predicted domain unit response rate and the target response rate are, respectively

$$URR_h^P = \frac{\left(r_{1h} + (m_{1h}^{i} * q_h^i)\right) + (m_{1h}^{e} * q_h^e * K_h)}{n_h}$$

$$URR^T \ = \frac{\sum_h \left(\left(r_{1h} + (m_{1h}^{i} * q_h^i)\right) + (m_{1h}^{e} * q_h^e * K)\right)}{\sum_h n_h}$$

We include the following two constraints to prevent oversampling in domains that are predicted to reach their target response rate and under sampling in domains that cannot reach their target response rate.

Any views expressed on statistical issues or operational procedures are those of the authors and not necessarily those of the U.S. Census Bureau

$$K_h = \begin{cases} 0, & \dfrac{r_{1h} + \left(m_{1h}^i * q_h^i\right)}{n_h} \geq URR^T \\ \infty, & \dfrac{r_{1h} + \left(m_{1h}^i * q_h^i\right) + (m_{1h}^e * q_h^e)}{n_h} < URR^T \end{cases}$$

Otherwise, for a given subsampling rate the predicted domain unit response rate must be greater than the target unit response rate; with the constraints that all $K_h$are bounded between zero and one (zero = no subsampling and one = full NRFU) and the subsample size is equal to $K * \sum_h m_{1h}^e$.

$$URR_h^P \geq URR^T$$

Any views expressed on statistical issues or operational procedures are those of the authors and not necessarily those of the U.S. Census Bureau