

# Investigating Adaptive Nonresponse Follow-up Strategies for Small Businesses through Embedded Experiments

Katherine Jenny Thompson<sup>1</sup> and Stephen J. Kaputa<sup>2</sup>

The U.S. Census Bureau is investigating adaptive Nonresponse Follow-Up (NRFU) strategies for single unit businesses in the 2017 Economic Census. These collection protocols require a suite of viable alternative procedures that can be implemented. With business surveys, the majority of cognitive research and nonresponse follow-up procedures focus on collection methods that obtain valid response data from the larger businesses, and there is relatively little quantitative or qualitative research for small businesses. Moreover, the contact methods for small businesses are often constrained by budget limitations. Business programs at the U.S. Census Bureau rely on mailed reminder letters and supplemental promotional materials, with options for certified and bulk mailings. To explore the benefits and disadvantages of the proposed alternative nonresponse follow-up procedures for small businesses, we conducted a field experiment embedded in the 2014 Annual Survey of Manufactures, an annual program that has similar data collection procedures and sampling units as the Economic Census. This article describes the study and presents the results, then discusses how the recommended nonresponse follow-up procedures are implemented in an adaptive collection design test presently being conducted in the 2015 Annual Survey of Manufactures.

*Key words:* Split panels; business surveys; adaptive collection design.

## 1. Introduction

The U.S. Census Bureau is investigating adaptive Nonresponse Follow-Up (NRFU) strategies for single unit businesses in the 2017 Economic Census. With an adaptive collection or response design, paradata and sample data are used to determine whether and when to change the current procedures (Schouten et al. 2013 and Groves and Heeringa 2006). Of course, for these strategies to work, the survey practitioner must have an available suite of viable alternative procedures to be implemented. For example, Groves and Heeringa (2006) discuss two-phase sample designs, where the first phase of collection surveys all sampled units using a relatively inexpensive protocol and the second phase of collection is a probability subsample of units who receive a more expensive but *provably effective* collection treatment.

<sup>1</sup> Economic Statistical Methods Division, U.S. Census Bureau, 4600 Silver Hill Road, Washington DC 20233, U.S.A. Email: Katherine.J.Thompson@census.gov

<sup>2</sup> Economic Statistical Methods Division, U.S. Census Bureau, 4600 Silver Hill Road, Washington DC 20233, U.S.A. Email: Stephen.Kaputa@census.gov

**Acknowledgments:** The authors thank Laura Bechtel, Cha-chi Fan, Carma Hogue, Eddie Salyers, the Associate Editor, and two anonymous referees for their useful comments on earlier versions of this article and Robert Struble for designing and implementing the panel selection procedures for the 2014 and 2015 ASM tests. Any views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

With business surveys, the majority of cognitive research and NRFU procedures focus on collection methods that obtain valid response data from the larger units (see [Willimack and Nichols 2010](#); [Snijkers et al. 2013](#); [Thompson and Oliver 2012](#); and [Thompson and Washington 2013](#), among others), attempting to control both measurement errors (making it easier for businesses to provide valid data under the same definitions as requested) and nonresponse errors. Why? Business populations are highly skewed, with a few sample units contributing to the majority of the industry totals. Consequently, business survey operational procedures are designed to increase the likelihood of obtaining valid responses from large units. Larger businesses are more likely to be contacted personally if there are questions about the data or as planned unit nonresponse follow-up. Account managers are assigned to maintain ongoing personal contact with the largest companies, especially those that are surveyed in many different programs ([Brady 2016](#)). Focus groups and business site interviews target large companies that operate in more than one industry, simultaneously attempting to gain insight into effective collection methods while recruiting survey response. In a sense, adaptive collection or (responsive) design practices are inherent in the contact strategy practices for large businesses surveyed by U.S. Census Bureau economic programs, especially in the NRFU procedures. Accordingly, the unit response rates and total quantity response rates (item-level) for the large businesses included with certainty (sampled with probability = 1) are often well above the 70% recommended by the [2006 Federal Register Notice](#) ([Knutson and Cepluch 2016](#); [Lineback and Fink 2012](#); [Thompson and Oliver 2012](#); [Thompson et al. 2015](#)).

In contrast, the small establishments receive very little personal contact (if any), and there is limited cognitive research on preferable contact strategies to draw upon. In general, smaller businesses are mailed reminder letters, but rarely receive telephone reminders or other personal contact. To summarize, unit nonresponse from large businesses is monitored and “managed,” whereas small business nonresponse tends to be downplayed. As response rates are decreasing in many programs, this small business subpopulation becomes more important in the survey totals and improvements in collection protocols are sought.

The Economic Census poses extra challenges with determining respondent contact strategies. This quinquennial program samples *establishments* (business locations), not companies (firms), collecting a core set of data items from each establishment called general statistics as well as information on the revenue obtained from product sales and other industry-specific variables. Establishments fall into two broad categories: (1) *Single unit establishments* own or operate a business at a single location and are classified into a single industry and (2) *Multi unit establishments* comprise of two or more establishments that are owned or operated by the same company. Multi unit companies receive one questionnaire per establishment and their packets may contain a variety of different forms depending upon the industries in which its establishments operate. Additional completeness procedures are implemented to ensure that the majority of a multi unit company’s questionnaires are received, as are additional data quality comparisons (examining aggregated establishment data to originating company administrative or auxiliary variables).

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 (2010) notes 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 (technically, Hedlin et al. (2008) provides evidence that larger businesses perceive surveys as being less burdensome than smaller businesses). In contrast, multi unit companies are more likely 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 2010).

The 2017 Economic Census will incorporate two major collection changes. First, data collection will be entirely electronic (web based) with no paper form alternatives; previous censuses have been primarily mail-out/mail-back and included a paper instrument. Second, the Economic Census will use the expanded North American Products Classification System (NAPCS), allowing cross-sector reporting instead of industry-specific product coding. These major changes are the catalyst for conducting embedded experiments on data collection features in several ongoing annual business surveys during the interim intercensal period. The results from these tests will be used to develop the census data collection strategies for 2017.

Our research investigates NRFU strategies designed to elicit response from single unit establishments. These strategies are field tested in the Annual Survey of Manufactures (ASM), selected for reasons provided below. In part, we are motivated by the suggested differences in collected data quality between large and small businesses described in Thompson and Washington (2013), Willimack and Nichols (2010), Bavdaž (2010), and Thompson et al. (2015). We are also motivated by the success reported in Kirgis and Lepkowski (2013), who describe how making specific changes to the data collection procedures to target specified “hard to reach” establishments improved estimates.

Returning to the theme of requiring alternative collection treatments for adaptive collection designs, 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. For example, personal phone calls and more frequent contacts have been shown to be effective for eliciting response from small businesses (Marquette et al. 2015), but are quite expensive and are generally not considered by the Census Bureau as financially viable options. Certified mailings have likewise been proven effective (Marquette et al. 2015). However, inexpensive procedures are generally preferred and are heavily utilized. These procedures include reminder letters, along with other supplementary printed materials.

The tone and content of these materials can affect response. Torres van Grinsven et al. (2014) categorizes businesses’ response motivation as *intrinsic* (“completely self-determined, internally rewarding motivation”) or as *extrinsic* (mandated by outside authorities). Material that highlights due dates and mandatory response emphasizes the extrinsic motivating forces, whereas material that describes aspects of survey utility appeal to more internalized motivating forces. These messages can be framed positively or negatively. Positively framed messages underscore the benefits of responding (Dillman et al. 1996). With a business survey, these materials might include a promotional message

illustrating the importance of the collected data to national estimates used for policy decisions (social utility motivation) or to explain how the published survey estimates can be used to benefit the designated establishment (egoistic motivation). Positively framed appeals have been shown to increase response rates with business surveys. [Hedlin et al. \(2008\)](#) describes an embedded experiment in a business survey conducted by Statistics Sweden that compared the effects of promotional material describing the uses and importance of collected data to material that offered free key estimates to the respondent at a later date. Neither treatment was shown to affect response rates, although there was a positive association between one of the incentive treatments and perception of the survey's usefulness. [Torres van Grinsven et al. \(2014, 592\)](#) report anecdotal findings from their cross sectional study of businesses, stating, "a friendly tone and language as an expression of a correct relationship seemed to be expected in communication that was addressed to businesses; they might have even been indispensable for survey requests to be considered." Alternatively, negative material carries a threatening message such as legal action for mandatory surveys. [Torres van Grinsven et al. \(2014\)](#) discuss detrimental effects on small business attitude and response when presented with negative or punitive reminder letters. [Tulp et al. \(1991\)](#) reports improved response rates by highlighting the legal obligation to respond to mandatory (business) surveys via a printed reminder on the survey notification envelope. Although these results are important, to our knowledge there are no studies that specifically contrast effects on response rates and respondent conditioning resulting from positively framed or negative framed material in a business survey setting as applied exclusively to small businesses.

As mentioned above, previous research conducted in the 2007 and 2012 Economic Censuses demonstrated that response rates for single unit establishments in low responding industries are improved by including a certified mailing in the NRFU protocol ([Marquette et al. 2015](#)). These studies were limited in scope and did not examine aspects of quality other than response rate. Our eventual goal is to implement a more adaptive collection protocol for all single unit businesses in the 2017 Economic Census, so we are interested in examining the effects of the proposed alternative NRFU procedures on *all* single unit establishments in a variety of industries. Consequently, we embedded a split panel study as field test in the 2014 ASM.

We selected the ASM as a testing ground 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 is a subset of the manufacturing sector's Economic Census questionnaire (the ASM is only conducted in non-census years) and the ASM uses the same editing and imputation procedures as the Economic Census. Since we are ultimately concerned with quality effects on collected items from small businesses in the Economic Census, the ASM is therefore an excellent testing ground. 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. Moreover, not all survey sponsors were comfortable with the risks associated with embedded experiments in stable ongoing programs.

Using a split panel design allows us to examine different aspects of response including response rates, timeliness, and potential nonresponse bias as characterized by respondent balance with respect to sample characteristics. These aspects address three of the five dimensions of data quality identified by Eurostat (see <http://ec.europa.eu/eurostat/web/income-and-living-conditions/quality>), namely *accuracy*, *timeliness*, and *coherence*. However, the Eurostat definition of accuracy refers to total survey error (i.e., sampling error + nonsampling error), which is difficult to measure directly. Instead, we assume that the originally drawn sample represents a complete cross-section of the target population, considering a response set to be “representative” of the sample when our collective set of measures for nonresponse bias do not provide evidence otherwise. These indicators are presented in Section 2.

For this experiment, all ASM single unit cases were randomly assigned to one of three separate panels, each receiving a different NRFU protocol (treatment). Multi unit establishments were excluded from this study as the ASM program managers have well-established and highly successful NRFU strategies for this subpopulation. Section 3 describes the study design, outlines the different treatments, and presents results. Section 4 presents the design and proposed analysis strategy for an ongoing embedded experiment in the 2015 ASM, which will couple nonrespondent subsampling of single unit establishments with the most effective follow-up procedures determined from the 2014 test. 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 Metrics

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 measurable among respondents). Wagner (2012) presents a useful typology for alternative indicators for the risk of nonresponse bias that incorporates this framework: (1) indicators involving the response indicator; (2) indicators involving the response indicator and frame data or paradata; and (3) indicators involving the response indicator, frame data or paradata, and the survey data. We classify the indicators described below using this typology.

Single unit establishments can vary greatly in size (e.g., total receipts, annual payroll, number of employees). The ASM is a stratified sample, and the largest single unit establishments in an industry are included with certainty. We use certainty status in our analyses as a proxy for size in all analyses.

### 2.1. Indicators Involving the Response Indicator: Type (1) and Type (2) Indicator Analyses

For business surveys, unit response rates are computed as unweighted ratios of respondents to eligible cases. This avoids overrepresentation of the smaller cases with larger weights in the response rate. For computation of the official rates, a respondent is

defined as 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). At the survey level, the response rates are Type 1 indicators. However, we are interested in treatment effects by nonresponse treatment panel (paradata) and by size category (frame data = certainty status) are therefore largely confined to Type 2 indicator analyses.

Examining plots of response rates over time by treatment panel (and size subdomain) can provide indications of potential treatment effects. We validate such observations by testing the hypothesis of independence of treatment and response rate using chi-squared tests for independence for the complete set of test cases and by certainty and noncertainty subdomain. Although the official unit response rates are computed without sampling weights, testing for differences without incorporating complex design features can lead to erroneous conclusions (Rao and Scott 1987). Using the Rao-Scott adjusted test implemented in PROC SURVEYFREQ addresses this problem (SAS/STAT(R) 9.3 User's Guide 2015), but incorporates the sampling weights, yielding different response rates than the official ones [Note: with our results, the differences in corresponding rates were trivial]. The SAS procedure incorporates the finite population correction (fpc) adjustment needed for the noncertainty subdomain to incorporate the reduction in sampling variance from without-replacement sample. Consequently, we conducted both unweighted and weighted analyses for the noncertainty case evaluations and unweighted analyses only for the certainty case evaluations. The unweighted analyses use the standard Pearson statistics; the weighted analyses use the Rao-Scott adjusted test.

Recall that we are interested in measuring timeliness as an aspect of data collection quality. It is possible that one or more of the treatments affects the length of response time, without impacting overall response. Time to respond can be modeled as failure-time data using available paradata and analyzed with survival analyses methods. We fit Cox proportional hazards regression models to the failure-time data, modeling the hazard function ( $\lambda_i(t)$ ) for each unit  $i$  at time  $t$  as  $\lambda_i(t) = \lambda(t; \mathbf{Z}_{i(t)}) = \lambda_0(t) \exp(\mathbf{Z}'_i(t)\beta)$ , where  $\mathbf{Z}_{i(t)}$  is a vector of explanatory variables at time  $t$  for unit  $i$  and  $\beta$  is the associated unknown vector of regression parameters, assumed to be the same for all individuals (Cox 1975). The primary statistic of interest is the hazard ratio: a value larger than one indicates a positive effect on response due to a treatment, whereas a value less than one is indicative of the converse. A secondary statistic of interest is the regression parameter, providing confirmatory evidence of the direction of the effect. Ideally, a beneficial NRFU protocol both positively affects response rate (increases it) and reduces length of time to respond.

To account for the complex survey design, we use PROC SURVEYPHREG to predict the onset of response and regress on treatments (SAS/STAT(R) 9.3 User's Guide 2015). As with the previous analysis, the fpc-adjustment is needed for the noncertainty subdomain, but cannot be used for the certainty subdomain (we use PROC PHREG).

Lastly, we are interested in examining the "representativeness" of the respondent composition of each treatment panel. The balance indicator and the distance indicator described in Särndal and Lundquist (2014) measure the degree to which the response set is similar to the full sample with respect to auxiliary variables or paradata available to all units on the frame.

Let

$y$  = characteristic of interest, subject to nonresponse

$x$  = auxiliary variable available for all sampled units

$P$  = weighted response rate =  $\sum_{i \in s} w_i I_i / \sum_{i \in s} w_i$ , where  $I_i$  is a unit response indicator and  $w_i$  is the design weight

Assume that  $y \approx \beta x + \varepsilon$ . Following [Särndal and Lundquist \(2014\)](#), define

**Balance**       $B_x = \bar{x}_r - \bar{x}_s$ , the difference between mean value for respondents and mean value for sampled units where  $\bar{x}_r = \sum_{i \in s} w_i x_i I_i / \sum_{i \in s} w_i I_i$  and  $\bar{x}_s = \sum_{i \in s} w_i x_i / \sum_{i \in s} w_i$ .  $B_x = 0$  is an indicator that the respondent set is a random sample of the parent sample for all collected variables correlated with  $x$ , the auxiliary variable.

**Imbalance**      Measured as  $IB_x = (\bar{x}_r - \bar{x}_s)' \Sigma_s^{-1} (\bar{x}_r - \bar{x}_s)$        $\Sigma_s = \sum_{i \in s} w_i x_i x_i' / \sum_{i \in s} w_i$ .

A *balance indicator for variable  $x$*  is given as  $BI_x = 1 - 2P\sqrt{IB_x}$ . This measure is bounded between 0 and 1, with values close to 1 indicating balance on the respondent sample for the studied variable. However, it tends to overestimate this quality.

**Distance**       $D_x = \bar{x}_r - \bar{x}_{nr}$  the difference between mean value for respondents and mean value for nonrespondents on variable  $x$ . This is measured as  $D_x = (\bar{x}_r - \bar{x}_{nr})' \Sigma_s^{-1} (\bar{x}_r - \bar{x}_{nr})$ .

Ideally, the balance indicator should be near 1 and distance indicator should be near 0. These measures were designed as process metrics to be monitored during data collection, similar to other process metrics such as R-indicators: see [Ouweland and Schouten \(2014\)](#) and [Schouten et al. \(2009\)](#), among others. The imbalance and distances measures ( $IB_x$  and  $D_x$ ) are approximate chi-square statistics with one degree of freedom and can therefore be tested for significant difference from zero using a large-sample approximation ([Johnson and Wichern 1988, 190–191](#)) with a few strong caveats, that is, the variance estimator is very conservative and does not incorporate the finite population sample design features. Likewise, contrasts between panels (within subdomain) can be tested using the multivariate Hottellings- $T^2$  analogue. However, these tests are extremely conservative, making it difficult to detect differences from zero (individual tests) or between panels (multivariate test).

## 2.2. Indicators Involving Survey Outcome: Type 3 Indicator Analysis

Type 3 indicators combine response rates, frame or auxiliary data available for all sampled units, and survey data to model potential effects of nonresponse bias. The proxy pattern-mixture (PPM) analysis first proposed by [Andridge and Little \(2011\)](#) falls into this indicator category. In brief, the PPM model reduces a set of fully observed auxiliary variables to a single “proxy” variable  $X$ . The joint distribution of a survey outcome  $Y$  and this proxy  $X$  is modeled as a bivariate normal distribution with separate parameters for respondents and nonrespondents (a pattern-mixture model). [Andridge and Thompson](#)

(2015b) develop a PPM model using a bivariate gamma model that is more appropriate for the studied skewed business populations. Either formulation produces adjusted estimates of the mean of  $Y$  under different missingness mechanisms, explicitly specified in the model used to link the proxy and outcome variable.

The Fraction of Missing Information (FMI) has been proposed as a metric for assessing the risk of nonresponse bias for a specific adjusted survey estimate (Wagner 2010, 2012; Andridge and Little 2011; Andridge and Thompson 2015a, 2015b). The FMI is a measure of loss of precision due to nonresponse, and is the ratio of between-imputation variance to total variance for a specific estimator (Little and Rubin 2002). The FMI value for a given  $Y$  is bounded between 0 and 1. In the PPM framework, FMI is computed with respect to an assumed response mechanism and a specified imputation model (in the application described in Section 3, a no-intercept regression model whose predictor is a frame variable), 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 compute 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 MNAR 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).

Wagner (2010) computes the FMI of several key variables during survey collection to study whether additional data collection decreases nonresponse bias effects over time. Our evaluation is analogous. We are particularly interested in seeing whether the increased response rates in a treatment panel have a corresponding beneficial effect in terms of nonresponse bias reduction on more than one variable. Furthermore, examining the FMI for an item by treatment panel while holding the prediction model constant provides insight into the composition of the respondent data population, building on the balance and distance indicators presented earlier by examining collected survey data. If the respondent set is “balanced,” we expect to see low FMI and we hope to see small differences in corresponding FMI estimates (MAR vs. NMAR).

The FMI values are strongly related to the strength of the predictors used in the proxy. In some instances, a treatment effect might be completely ameliorated by an excellent predictor (strong proxy). The converse can also be true if the relationship between predictors and outcome is not strong (weak proxy).

### **3. 2014 Panel Test: Testing Alternative NFRU Procedures on Small Businesses**

#### *3.1. Background and Study Design*

Section 1 outlines the motivation for selecting the ASM as the vehicle for testing NRFU strategies for single unit establishments for the upcoming 2017 Economic Census. However, there are some key differences in survey design between the two programs. The term “economic census” is a slight misnomer since the majority of sectors select a small stratified sample of single unit establishments. The ASM is a stratified sample, with strata

are defined by six-digit industry code using the North American Industry Classification System (NAICS). The industry strata are further subdivided into two subdomains: certainty and noncertainty. The largest multi and single unit establishments in an industry are included with certainty. The remaining establishments are a stratified Pareto-PPS sample, selected with probability proportional to a composite measure of size (MOS) based on annual payroll from the most recent Economic Census and realized sampling weights ranging from 1.01 to 20. The ASM surveys approximately 50,000 establishments selected from a universe of about 328,500. About 7,000 of the approximate 20,000 establishments included with certainty are single unit establishments, and about 10,000 of the approximate 30,000 noncertainty are single unit establishments. To reduce respondent burden, units below a certain size threshold do not receive a questionnaire; their data are obtained using administrative records and model imputation. Similarly, the ASM imputes complete records for unit nonrespondents. See <http://www.census.gov/manufacturing/asm/> for additional information on the ASM methodology.

For both the ASM and the Economic Census, the collection design varies by type of unit, with contact strategies designed to ensure that the largest cases provide valid response data. Nonresponding multi unit establishments and selected larger single unit establishments receive more frequent NRFU contacts and can include more expensive personal contacts, such as phone follow-up. The remaining small single unit establishments receive reminders, but are very unlikely to receive personal contact. The contact strategy for single unit establishments in the 2014 ASM relied entirely on mail outreach. Altogether, there are five possible contacts:

- Initial contact letter (all sampled units), providing a deadline and requesting internet response via a secure system that includes username and password
- First NRFU: reminder letter stating that response is past due and again requesting response via internet.
- Second NRFU: reminder letter plus paper questionnaire. Historically, the questionnaire was included in the initial mailing; this is the first time the questionnaire mailing has been delayed until the second NRFU. This was done primarily to push internet collection over paper collection, but also to save money.
- Third and fourth NRFU: progressively threatening letters requesting response. See [Figure 1](#) the [Appendix](#) for a sample.

Historically, the Economic Census and the ASM have been mail-out/mail-back collections. Although both programs are moving towards complete internet collection, there were concerns that not all businesses could respond through this mode, coupled with concerns about previous respondent conditioning towards paper collection. Especially because of the latter, the ASM subject matter experts were not comfortable with a complete push towards internet collection for the 2014 survey year. Instead, they preferred to “transition” respondents by including a paper questionnaire at some stage of the NRFU and then dropping the paper instrument entirely from the 2015 collection onward.

Three separate NRFU strategies were tested in the 2014 embedded experiment. To obtain the treatment panels, the ASM single unit sample was blocked on three-digit NAICS industry and certainty/noncertainty subdomain. Establishments were sorted within blocks by their sampling frame MOS (see Subsection 3.1; hereafter, we refer to this

Table 1. Contact strategies by treatment panel.

Panel	Initial mail	First NRFU	Second NRFU	Third NRFU	Fourth NRFU
Control (C)	Letter	Letter	Form	Letter	Letter
Treatment 1 (T1)	Letter	Letter	<i>Certified Letter</i>	<i>Form</i>	Letter
Treatment 2 (T2)	Letter	Letter	<i>Letter/Negative Flyer</i>	<i>Form</i>	Letter

variable as MOS) and systematically assigned to treatment panels using a random start. Thus, each treatment panel's composition is balanced by industry, certainty/noncertainty subdomain, and establishment size. At the time of initial mail, each panel contained approximately 5,700 single units and was expected to have 3,300 nonrespondents at the time of the second NRFU.

Table 1 presents the tested contact strategies by treatment panel. Each round of contact followed the same fixed calendar schedule. The experimental treatments (NRFU strategies) that do not follow the normal contact strategy are italicized. For all panels, the initial contact and first NRFU procedure are the same. The control panel uses the contact strategies for the 2014 ASM described above. Treatment panels 1 and 2 respectively test the effectiveness of a certified letter (T1) and the negative flyer/letter provided in Figure 2 of the Appendix (T2) as the second NRFU. All NRFU protocols include a questionnaire mailing (denoted as form in Table 1); however, this mailing is delayed until the third NRFU contact for the T1 and T2 panels. Lastly, all treatment panels received the same final letter as fourth NRFU contact.

### 3.2. Results

A major benefit of a successful adaptive collection strategy would be to reduce the nonresponse bias in the survey estimates [Note: it can be argued that cost savings are the more compelling factor, and this must be considered in the final design]. In our experimental setting, it is possible that a given NRFU strategy could improve response rates without resulting in any detectable differences in the estimate quality if all three sets of respondent samples are "representative" (i.e., all sample units' response mechanism is missing completely at random or missing at random) or are all equally lacking in a subdomain (e.g., the smallest establishments). On the other hand, it is not difficult to determine a single "treatment effect" when one NRFU method exhibits improved performance on one measure of nonresponse bias over the others.

#### 3.2.1. Type (1) and Type (2) Indicator Analyses

Figure 1 separately plots "proxy" unit response rates over time by treatment panels for certainty cases and for noncertainty cases within panel. These are not the officially released unit response rates, as the study data had undergone minimal editing and imputation at the time of analysis. We categorized a unit as responding if it provided a value for annual payroll; this value may or may not be used in the final tabulations (an imputed or administrative data value may be substituted). For simplicity, hereafter we refer to this proxy rate as the unit response rate.

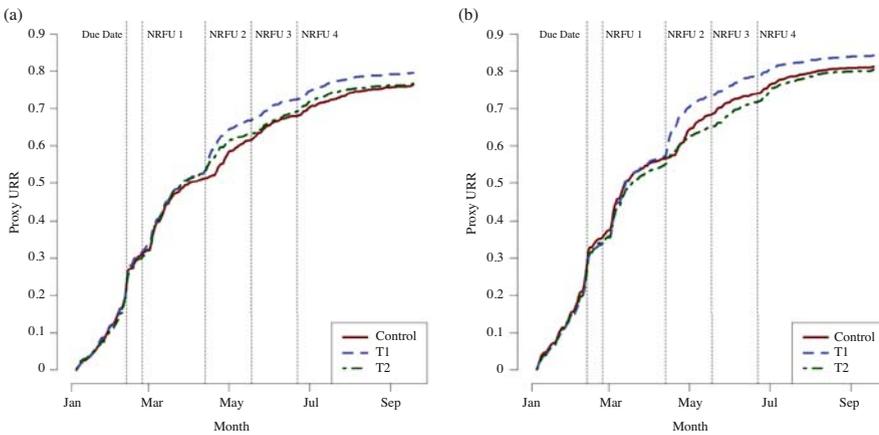


Fig. 1. Unit response rates by treatment panel certainty (a) and noncertainty (b).

Regardless of certainty status, these plots provide strong visual evidence of an improved response rate when the certified letter is used for the second NRFU attempt, in contrast to the other treatments, which have virtually indistinguishable response rate plots. We substantiate this observation using chi-square tests of independence of treatment and response rate whose results are presented in Table 2.

All tests reject the null hypothesis (at  $\alpha = 0.05$ ), providing evidences that at least one treatment results in a different unweighted response rate, with an “across the board” effect (i.e., not confined to a single subdomain). However, the first round of NRFU is the same for all treatments, and it is possible that the differences in unit response rates could be attributable to a pre-existing difference in unit response rates between panels that existed before the second round of follow-up. To assess this, we performed similar analyses within subdomain, treating the cases that responded *before* the second round of NRFU as the respondents. None of these tests provide sufficient evidence (at  $\alpha = 0.05$ ) of a “treatment panel effect” on the unit response rate before the second follow-up procedure is introduced. It does appear that the difference in response rates between treatment panels is primarily attributable to the differences in NRFU procedures at the second round of NRFU.

Table 2. Tests for equivalence of unit response rate by treatment.

Treatment panel		Certainty		Noncertainty	
		Unweighted	Unweighted*	Unweighted*	Weighted**
C	Response rates	69.80	78.90	76.97	
T1		73.21	81.73	80.86	
T2		69.57	78.10	75.85	
Test statistic		9.50	15.05	11.48	
p-value		0.0087	0.0005	0.0030	

\* $\chi^2(2)$ , Pearson test.

\*\* $\chi^2_{RS}(2)$ , Rao-Scott Adjusted test.

Table 3. Proportional hazards regression model tests by subdomain.

Parameter	Certainty		Noncertainty	
	Estimate	Hazard ratio	Estimate	Hazard ratio
T1 (certified letter – form)	0.06	1.07	0.09*	1.10*
T2 (letter/flyer – form)	0.08*	1.08*	–0.03	0.97
Contrast (T1-T2) test $p$ -value	–0.36		<0.0001	
Global test $p$ -value	0.0542		<0.0001	

Inspecting Table 2, it appears that the unit response rates for the control panel (C) and the letter/flyer (T2) panels are very similar. Performing reduced tests – dropping the T1 treatment panel – yields no treatment effect for all single unit cases, providing further evidence of no treatment effect using the negative letter (unweighted  $\chi^2(1) = 0.48$ ,  $p$ -value = 0.49; weighted  $\chi_{RS}^2(1) = 0.44$ ,  $p$ -value = 0.51), for certainty cases (unweighted  $\chi^2(1) = 0.02$ ,  $p$ -value = 0.87), and for noncertainty cases (unweighted  $\chi^2(1) = 0.62$ ,  $p$ -value = 0.42; weighted  $\chi_{RS}^2(1) = 0.49$ ,  $p$ -value = 0.49). Ultimately, these results provide evidence that sending a certified letter, followed by mailing the form to the remaining nonrespondents, increases the response rate over the other two treatments. Equally important, these results provide evidence that sending a strongly worded message of delinquency and obligation, followed by mailing the form to the remaining nonrespondents does not improve the response rate.

We examine failure-time as a function of treatment panel (baseline = control panel). Table 3 provides the maximum likelihood estimates for each parameter along with the associated hazard ratio. Statistically significant regression parameters and hazard ratios at  $\alpha = 0.05$  are indicated by an asterisk.

For the *noncertainty* single unit cases, these results provide evidence of a positive effect on time to respond for the cases in the certified letter treatment (T1) panel and an increased probability of responding over the current procedure. The increased probability of responding translates to an increased overall response rate in this subdomain, as shown above. These smaller units are responding more quickly and at a higher rate, providing convincing evidence that the certified letter treatment is more effective than the current procedure for eliciting responses from the noncertainty single unit cases. Further confirmation is provided by the pairwise contrast test of the T1 and T2 treatments in the noncertainty population ( $p$ -value = 0.0001).

For the *certainty* single unit cases, there is evidence of a positive effect on time to respond for the cases in the letter/flyer treatment (T1) panel and an increased probability of responding over the current procedure. Recall, however, that there is no significant difference in unit response rate for these units in the control and T2 panels. Thus, the increased probability of eventually responding has no practical effects. Indeed, it appears that responding units are simply answering the survey in a timelier manner, with a nonsignificant contribution from “new” units being converted to respondents via the letter/flyer treatment.

The previous analyses provided evidence that the T1 NRFU strategy elicits more responses from small noncertainty units than the other two NRFU strategies under consideration. If this is the case – and the other two panels *underrepresent* these smaller

Table 4. Distance and balance measures by panel on administrative payroll and MOS.

Treatment	Certainty				Noncertainty			
	Payroll		MOS		Payroll		MOS	
	BI <sub>x</sub>	D <sub>x</sub>						
C (control)	0.98	0.00	1.00	0.00	0.97	0.01	0.98	0.00
T1 (certified letter – form)	0.98	0.00	0.97	0.01	1.00	0.00	0.99	0.00
T2 (letter/flyer – form)	1.00	0.00	1.00	0.00	0.99	0.00	1.00	0.00

units in the response set – then the T1 sample composition should be a more “representative” subsample of the ASM single unit sample. Table 4 presents the distance and balance measures on 2014 administrative payroll and MOS by panel and certainty status (subdomain).

For this analysis, we are interested in whether one particular treatment appears to yield a more representative sample than the others or if all have equally balanced or equally unbalanced samples. Overall, all panels are balanced with respect to both auxiliary variables. In fact, there were no significant differences from zero for any individual *imbalance* measure (all  $p$ -values greater than 0.99), and the multivariate panel contrast tests are likewise not significant. The corresponding distance measures are very close to zero, with none of the conservative individual chi-square tests detecting a difference ( $p$ -values ranging from 0.63 to 0.99) and nonsignificant differences between panels. These measures were originally proposed as process control measures. In that spirit, we make a few observations. First, for the certainty units, the T2 panel (letter/flyer – form) is perfectly balanced on both auxiliary variables and has the minimal difference (at the fourth decimal place, not shown). For the larger single unit establishments, this provides some support for the subject matter experts’ contention that the harsh tone of the flyer elicits response at higher rate than the other NRFU protocols. This is not the case for the noncertainty units. Here, the two alternative treatments (T1 and T2) result in *apparently* more balanced respondent samples than that obtained with the current (C) procedure. One could argue that the marginally improved results on 2014 administrative payroll with the T1 NRFU procedure are more relevant than the reverse seen with the T2 NRFU procedure on the T2 sample because the administrative data are obtained from the concurrent collection period. Again, this is not a strong argument – or completely convincing evidence of superior balance for either treatment – given the optimality of the balance and distance indicators for all treatments and panels.

### 3.2.2. Type (3) Indicators

Using the bivariate-gamma PPM formulation (Andridge and Thompson 2015b), we compare the FMI within treatment panel on three separate survey items (payroll, employment, and receipts), producing a separate proxy for each by regressing the outcome variable on MOS within three-digit industry (a no-intercept linear regression model). As recommended by Andridge and Thompson (2015a, 2015b), we use multiple imputation to produce all estimates, with 200 draws given a burn-in period of 500 draw and thinning at every tenth draw. Our three study items have varying proxy strength: the payroll proxy fit

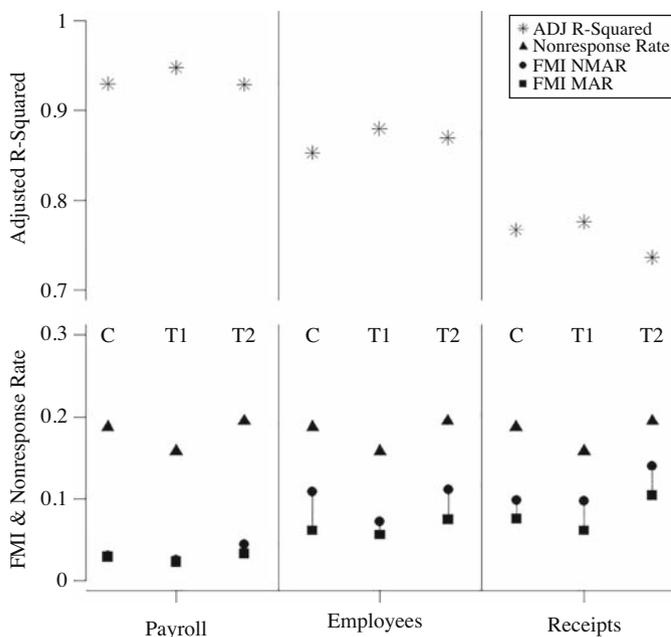


Fig. 2. FMI for noncertainty single units for three key survey variables by treatment panel under two alternative response mechanisms.

is extremely strong (as expected), the employment proxy is very strong, and the receipts proxy fit is weak.

Figure 2 presents the adjusted- $R^2$ , nonresponse rate, and FMI values by treatment panel for the three studied items, focusing exclusively on single unit noncertainty cases; there are insufficient data in the certainty subdomains to perform similar analysis and the combined analysis results does not differ from the noncertainty analysis. The models used to develop the payroll and employment proxies use all respondent data; five outliers were removed from the total receipts model to improve the fit [Note: these outliers would not be present in fully edited data].

Regardless of outcome variable, Figure 2 shows that all FMI values for all variables are well below the nonresponse rate, signaling that the studied estimates are not overly subject to nonresponse bias in any given panel (Andridge and Little 2011). However, the T1 panel estimates always have the lowest FMI of the three corresponding panel estimates under the MAR and NMAR response mechanisms. Since the regression models differ only in the treatment variable, this provides limited evidence that the T1 respondent data set is more representative on the studied variables than the others given the assumed models. Furthermore, with the exception of receipts, the T1 FMIs exhibit the smallest spread. With payroll, the small spread is likely a consequence of the extremely strong proxy. However, with employees, the reduced spread shows the combined effect of the fairly strong proxy and the small distance between respondents and nonrespondents. With receipts, the T1 FMI levels remain smaller than the T2 levels; the similar spread shows the effect of the weaker proxy on correcting nonresponse.

Not surprisingly, the FMI results for the payroll model are essentially identical to the distance and balance indicator results provided in Subsection 3.2.1., given that the MOS variable is largely derived from Economic Census payroll values. Restricting this analysis to these balance and distance indicators – or to the FMI from the payroll proxy – appears to be overly optimistic about the effect of NRFU protocol on respondent sample composition. Examining the FMIs of the employment and receipts proxies is more revealing, as they demonstrate a negative effect on the respondent sample for other – more difficult to impute – items obtained using T2 NRFU protocol.

Examined collectively and quantitatively, this embedded experiment showed:

- A NRFU strategy that includes a contact strategy that emphasizes the *value* of response from a sampled respondent such as a certified letter elicits responses from the smallest businesses in the sample. This is consistent with the results presented in [Torres van Grinsven et al. \(2014\)](#).
- A NRFU strategy that elicits response through negative reinforcement (stressing dereliction, for example) has no effect on the small business's response propensities. It does appear to decrease the response time for the larger surveyed companies, but did not increase the number of respondents.
- The increased response from sampled smaller businesses obtained via the T1 (certified letter) NRFU protocol improved the overall representativeness of the response set over the other considered methods. This effect was more apparent for items with lower item response-rates and weaker imputation models.

The positive effects on response rate and response composition are not entirely confined to this embedded experiment; similar findings with certified letter mailings in selected industries from the 2007 and 2012 Economic Censuses are presented in [Marquette et al. \(2015\)](#). Emphasizing the importance of the individual business' contribution to the survey has been shown to positively affect response in other business programs as well ([Kennedy and Phipps 1995](#)), without demonstrating a strong association between the solicited response and the resultant data quality, however. A limitation of our study is that we do not address measurement errors, which could increase the nonsampling error component of the total survey error from nonrespondent conversion, as pointed out by a referee. Nevertheless, the strong collective experimental evidence from this study has motivated a NRFU protocol change for single unit establishments beginning with the 2015 ASM collection. Henceforth, *all* nonrespondent single unit establishments will receive a certified letter reminder as the second NRFU contact.

#### 4. Ongoing 2015 ASM Test

The primary purpose of the 2014 embedded experiment was to determine a NRFU protocol that elicited improved response from small businesses sampled in the ASM. The larger research question is how to effectively implement this protocol in an adaptive collection design, balancing competing interests of costs and quality. One considered approach is the selection of a probability subsample of small single unit businesses for NRFU; an alternative approach is to continue the current procedure of NRFU of all originally sampled units, but target the more expensive procedures to subdomains that

have lower initial response. The 2015 ASM test considers both approaches, serving two purposes:

1. Compare quality effects of using targeted selection of nonrespondents to receive certified mail reminder compared to sending all nonrespondents a certified mail reminder letter (adaptive design NRFU protocol versus fixed design NRFU protocol)
2. Compare quality effects of selecting a probability subsample of nonrespondents for NRFU

As with the 2014 test, the target population will be the ASM single unit nonrespondents. The experimental design is different, however. All ASM noncertainty single unit cases receive the same initial contact letter, due date reminder letter, and first NRFU letter. This maximizes the usage of previously-proven contact strategies for this survey. After the first NRFU concludes, the ASM industries will be split into two separate panels, based on the blocking criteria such as sample size, proxy unit response rates, and total stratum payroll. All nonrespondent units in the control panel receive a certified letter reminder (second NRFU) and a strongly-worded reminder letter if they have not responded to the second NRFU attempt.

The treatment panel assignment is more complex. We use the optimized allocation described in [Kaputa et al. \(2014\)](#) to select a targeted systematic probability subsample of units in each industry (T1). These sampled units (T1) receive a certified letter reminder (second NRFU) and an OGC letter if necessary. The remaining complimentary units (T2) receive a reminder letter (not certified) and an OGC letter if necessary. Consequently, *all* nonresponding units in the treatment panel receive *some* form of NRFU. This approach is not dissimilar in spirit from the score-prioritization strategy presented in [Berthelot and Latouche \(1993\)](#). The earlier work proposed limiting expensive NRFU strategies to pre-selected units with a high expected impact on survey totals, but retaining a (less expensive) NRFU procedure for all total nonrespondents “to ensure that at least the status of the respondent (active, inactive, or out of scope) is obtained.” ([Berthelot and Latouche 1993](#), 419).

[Figure 3](#) illustrates the experimental design implemented in the 2015 ASM Adaptive Design Test.

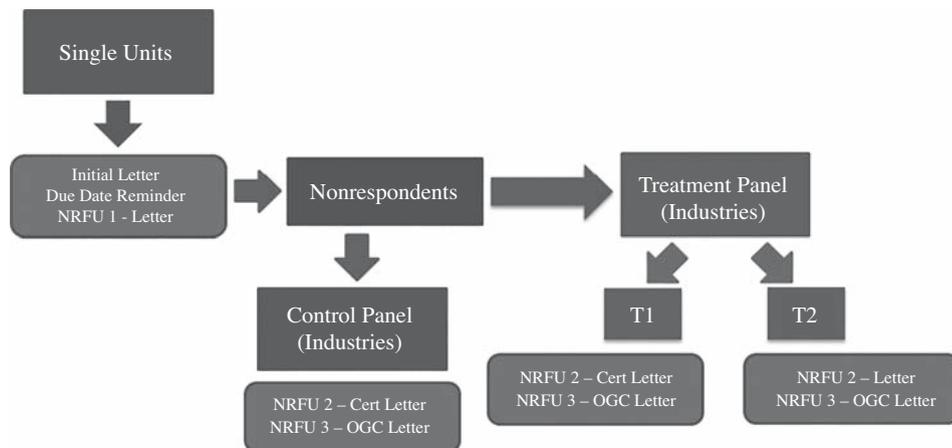


Fig. 3. Experimental design for 2015 ASM adaptive design test.

By retaining the subsampling parameters associated with the T1 selection, comparisons can be made between the Control panel and the T1 (subsampled units) panel, as well as the Control panel and the Test (T1 + T2) panel. The surveying costs with the Test panel will be slightly less than 50% of the Control panel costs, as the nonresponse sample allocation uses an overall 1-in-2 subsampling rate. Retaining the subsampling weights allows us to simulate the effects on survey estimates of using only a probability sample of nonrespondents. This experiment was completed in fall 2016 and the results are presently being analyzed.

## 5. Conclusion

Methodologists frequently rely on simulation studies to assess competing research proposals. This can work well when the underlying conditions are constant, for example, comparing alternative sampling designs on the same frame or developing an allocation strategy for nonrespondent subsampling. The ultimate objective of our research is to develop an adaptive collection strategy for the Economic Census that simultaneously reduces costs, maintains response quality, and allows timely and complete estimation. At surface glance, the originally proposed two-phase sample design appeared to be quite reasonable. However, having a toolkit of alternative follow-up procedures, with quantitative evidence on the performance of each is a key component of this approach – whether only subsampled units receive follow-up or all units receive some form of follow-up with the more expensive/effective protocol confined to a subsample.

Unfortunately, we did not have such a toolkit. Subject matter experts had anecdotal opinions (all varying), but no quantitative results. There is very limited literature on nonresponse follow-up strategies for small businesses. Cognitive research that relies on focus groups and business site interview with large companies could be useful, but might not necessarily extrapolate to similar sized establishments operating in different sectors, let alone to single unit establishments. In short, real experimentation – on the actual target audience – was required.

The presented case study proves the value of an embedded experiment in this situation. The test was not difficult to conduct and provided convincing evidence to the subject matter experts. The split panel design allowed for analysis on a variety of diverse small businesses in different industries, as opposed to focus groups. The lessons learned in administering the split panel test helped ensure a smooth operation in the subsequent (ongoing) planned experiment, and the program's managers have invited us to continue embedding experiments in future collection years. Our statistical analyses were compromised somewhat by the challenges of complex survey design effects, and ongoing production (changes in collected data throughout collection, especially due to late response). The power of the tests may be likewise detrimentally affected; especially the conservative tests performed in the noncertainty subdomains. Nevertheless, the analyses presented here provided convincing results that we believe can be extrapolated to other similar populations. Best of all, these analyses go beyond simple response rate comparisons and this framework can be applied in general to other embedded experiments.

## Appendix



UNITED STATES DEPARTMENT OF COMMERCE  
Economics and Statistics Administration  
U. S. Census Bureau  
Washington, DC 20233-0001

This is an official request from the United States Government. The Office of Management and Budget (OMB) approval number for these collections are 06070444, which expires 5/30/2016 and 0607 0440, which expires 4/30/2015.

**A Message from the Director, U. S. Census Bureau:**

Your 2014 Report of Organization, NC-99001, and 2014 Annual Survey of Manufactures, MA-10000(L) are overdue. These surveys are mandatory and were due by February 12, 2015.

Please report within 10 days. To begin, go to the website noted below, click the "Report Now" button, and enter your user ID and password.

Website:

User ID:

Password:

If you recently submitted your surveys, please visit the website noted above and use the "Self Service Log in" to verify your filing status.

**Please Note:** The U.S. Census Bureau is moving toward all-electronic collections. In addition to reducing our costs, the benefits to reporting electronically include:

- Built-in calculation features
- Edits that identify potential reporting issues
- Immediate confirmation of submission

The Report of Organization is used to update the Census Bureau's list of businesses, and it provides key source data for the County Business Patterns program and other statistical series. Companies, business analysts, and trade associations use this information in planning investments, production, and marketing.

The Annual Survey of Manufactures collects key data about the manufacturing sector of the nation's economy. These data are used to develop accurate estimates of domestic output and productivity, and they serve as a valuable resource for making sound decisions on economic trade policies.

Title 13, United States Code, Sections 182 and 224, requires your response and Section 9 guarantees that your response is confidential and will be used for statistical purposes only. Applicable provisions of the law are provided on the website noted above. You may use reasonable estimates if book figures are not readily available.

If you need assistance beyond what is available on our website, please contact us at 1-800-233-6136 Monday through Friday, 8:00 a.m. to 4:30 p.m. Eastern Time.

Thank you for your cooperation.

United States™  
**Census**  
Bureau  
Economic Statistics

Fig. 1. NRFU reminder letter (standard).



UNITED STATES DEPARTMENT OF COMMERCE  
Economics and Statistics Administration  
U. S. Census Bureau

## 2014 Annual Survey of Manufactures

### Your response is overdue:

Your MA-10000 report was due February 12, 2015.

Within the next 10 days, please visit us online to begin reporting. Refer to the enclosed letter for our secure website and your log in credentials. If you cannot report at this time, please go online and request an extension so we know when to expect your report.

### Your participation is required by law:

This survey is mandatory under an Act of Congress. Title 13, United States Code, Sections 182 and 224, requires your response. Section 9 guarantees that your response is confidential and will be used for statistical purposes only.

### Your data are critical:

Your manufacturing business is part of a scientifically selected sample and represents other business in the country of similar size and industry. Therefore, it is critical to receive your data to ensure we can publish reliable national estimates on current U.S. manufacturing industry outputs, inputs, and operating status.

Fig. 2. NRFU "Flyer" (T2 Panel).

## 6. References

- Andridge, R.R. and R.J.A. Little. 2011. "Proxy Pattern-Mixture Analysis for Survey Nonresponse." *Journal of Official Statistics* 27: 153–180.
- 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: <http://dx.doi.org/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: <http://dx.doi.org/10.1214/15-AOAS878>.

- Bavdaž, M. 2010. "The Multidimensional Integral Business Survey Response Model." *Survey Methodology* 36: 81–93.
- Berthelot, J.M. and M. Latouche. 1993. "Improving the Efficiency of Data Collection: A Generic Respondent Follow-up Strategy for Economic Surveys." *Journal of Business and Economic Statistics* 11(4): 417–424. Doi: <http://dx.doi.org/10.2307/1391632>.
- 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.
- Cox, D.R. 1975. "Partial Likelihood." *Biometrika* 62: 269–276. Doi: <http://dx.doi.org/10.2307/2335362>.
- Dillman, D.A., E. Singer, J.R. Clark, and J.B. Treat. 1996. "Effects of Benefits Appeals, Mandatory Appeals, and Variations in Statements on Confidentiality on Completion Rates for Census Questionnaires." *Public Opinion Quarterly* 60(3): 376–389.
- 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 (Statistics in Society)* 169: 439–457. Doi: <http://dx.doi.org/10.1111/j.1467-985X.2006.00423.x>.
- Federal Register Notice. 2006. OMB Standards and Guidelines for Statistical Surveys. Available at: [https://unstats.un.org/unsd/dnss/docs-nqaf/USA\\_standards\\_stat\\_surveys.pdf](https://unstats.un.org/unsd/dnss/docs-nqaf/USA_standards_stat_surveys.pdf) (accessed July 2017).
- 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.
- Johnson, R.A. and D.W. Wichern. 1988. *Applied Multivariate Statistical Analysis (2<sup>nd</sup> Edition)*. New Jersey: Prentice Hall.
- 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).
- Kennedy, J. and P. Phipps. 1995. "Respondent Motivation, Response Burden, and Data Quality in the Survey of Employer-provided training." In *Proceedings of the Annual Meeting of the American Association for Public Opinion Research*, May 1995, Ft. Lauderdale, FL. Available at: <http://www.bls.gov/osmr/pdf/st950250.pdf> (accessed September 2016).
- Kirgis, N. and J. Lepkowski. 2013. "Design and Management Strategies for Paradata-Driven Responsive Design: Illustrations for the 2006–2010 National Survey of Family Growth." In *Improving Surveys with Paradata*, edited by Frauke Kreuter. New York: Wiley.
- 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, August 4, 2016. Alexandria, VA: American Statistical Association.
- 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](http://www.amstat.org/meetings/ices/2012/papers/302146.pdf) (accessed February 2017).
- Little, R.J.A. and D.B. Rubin. 2002. *Statistical Analysis with Missing Data (2<sup>nd</sup> Edition)*. New York: Wiley.
- 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.
- Ouwehand, P. and B. Schouten. 2014. "Measuring Representativeness of Short-Term Business Statistics." *Journal of Official Statistics* 30(4): 623–649. Doi: <http://dx.doi.org/10.2478/jos-2014-0041>.
- 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.
- 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. Doi: <https://doi.org/10.1093/jssam/smu014>.
- "SAS/STAT(R) 9.3 User's Guide". *SAS/STAT(R) 9.3 User's Guide*. N.p., n.d. Web. 09 Oct. 2015.
- Schouten, B., M. Calinescu, and A. Luiten. 2013. "Optimizing Quality of Response through Adaptive Survey Designs." *Survey Methodology* 39(2): 29–58.
- Schouten, B., F. Cobben, and J. Bethlehem. 2009. "Indicators of Representativeness of Survey Nonresponse." *Survey Methodology* 35: 101–113.
- Snijders, G., G. Haraldsen, J. Jones, and D.K. Willimack. 2013. *Designing and Conducting Business Surveys*. New York: Wiley.
- 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.
- 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: <http://dx.doi.org/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*. Available at: <http://surveyinsights.org/?p=2991>.
- Torres van Grinsven, V., I. Bolko, and M. Bavdaž. 2014. "In Search of Motivation for the Business Survey Response Task." *Journal of Official Statistics* 30(4): 579–606. Doi: <https://doi.org/10.2478/jos-2014-0039>.
- Tulp, D.R., Jr., C.E. Hoy, G.L. Kusch, and S.L. Cole. 1991. "Nonresponse under Mandatory versus Voluntary Reporting in the 1989 Survey of Pollution Abatement Costs and Expenditures." In *Proceedings of the Section on Survey Research Methods*, August 1991. Alexandria, VA: American Statistical Association. Available at: <http://ww2.amstat.org/sections/srms/Proceedings/> (accessed February 2017).
- 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: <http://dx.doi.org/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: <http://dx.doi.org/10.1093/poq/nfs032>.

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

Received January 2016

Revised February 2017

Accepted March 2017