Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Strategies for subsampling nonrespondents for economic programs

by Katherine Jenny Thompson, Stephen Kaputa, and Laura Bechtel

Release date: June 21, 2018

Statistics Canada Statistique Canada

Canadä

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at STATCAN.infostats-infostats.STATCAN@canada.ca

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-877-287-4369

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard table symbols

The following symbols are used in Statistics Canada publications:

. not available for any reference period
.. not available for a specific reference period
... not applicable
0 true zero or a value rounded to zero
0^s value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p} preliminary
^{r} revised
x suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E} use with caution
F too unreliable to be published
* significantly different from reference category ($p < 0.05$)

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, June 2018
Vol. 44, No. 1, pp. 75-99
Statistics Canada, Catalogue No. 12-001-X

Strategies for subsampling nonrespondents for economic programs

Katherine Jenny Thompson, Stephen Kaputa, and Laura Bechtel[1]

Abstract

The U.S. Census Bureau is investigating nonrespondent subsampling strategies for usage in the 2017 Economic Census. Design constraints include a mandated lower bound on the unit response rate, along with targeted industry-specific response rates. This paper presents research on allocation procedures for subsampling nonrespondents, conditional on the subsampling being systematic. We consider two approaches: (1) equal-probability sampling and (2) optimized allocation with constraints on unit response rates and sample size with the objective of selecting larger samples in industries that have initially lower response rates. We present a simulation study that examines the relative bias and mean squared error for the proposed allocations, assessing each procedure's sensitivity to the size of the subsample, the response propensities, and the estimation procedure.

Key Words: Quadratic program; Unit response rate; Nonresponse adjustment; Systematic sampling; Optimal allocation; Two-phase sampling.

1 Introduction

Many federal programs are simultaneously experiencing declining response rates and reductions in funding. At the same time, these programs are required to maintain predetermined reliability levels and are encouraged to collect an increased number of data items and to publish more statistics. Of course, as nonresponse increases, the precision of the survey estimates will decrease from the original design levels and can be sensitive to nonresponse bias. Consequently, many federal agencies are investigating adaptive collection design strategies, where the term "collection design" refers to protocol(s) for collecting data.

With business surveys, the collection design may vary by type of unit. These populations are generally highly skewed; the majority of a tabulated total in a given industry is often provided by a small number of large businesses. Because publication statistics are generally industry totals or percentage change, missing data from the largest cases can induce substantive nonresponse bias in the totals, whereas missing data from the smaller cases (even those with large sampling weights) often have little *apparent* effect on the tabulated levels (Thompson and Washington, 2013). Thus, the contact strategies are designed to ensure that the largest cases provide valid response data. Figure 1.1 illustrates nonresponse follow-up (NRFU) procedures that differ by a survey-specific unit size classification, where both collection designs have fixed calendar schedules and a fixed NRFU budget.

For the large unit category, the NRFU procedures become progressively more costly (per unit) with the exception of the final contact attempt. In contrast, with the smaller units, the NRFU procedures do not include personal contact and are therefore less expensive.

Selecting a probability subsample of nonrespondents is a strategic feature of many responsive and adaptive collection designs (Tourangeau, Brick, Lohr and Li, 2016). Of course, this is not a new practice

1. Katherine Jenny Thompson, Economic Statistical Methods Division, U.S. Census Bureau, Washington, DC 20233. E-mail: Katherine.J.Thompson@census.gov; Stephen Kaputa, Economic Statistical Methods Division, U.S. Census Bureau, Washington, DC 20233. E-mail: Stephen.Kaputa@census.gov; Laura Bechtel, Economic Statistical Methods Division, U.S. Census Bureau, Washington, DC 20233. E-mail: Laura.Bechtel@census.gov.

for surveys. Indeed, nonrespondent subsampling has been a survey practice since first discussed in Hansen and Hurwitz (1946). Actually, the setting of the two-phase sample approach presented in Hansen and Hurwitz (1946) paper is quite similar to the business survey setting discussed here: an "inexpensive" mailed questionnaire to all sampled units (c.f. the "21st century design" that mails a letter containing a URL, user name, and password), followed by "expensive" personal interviews of subsampled nonrespondents (c.f. personal phone calls or certified reminder letters). Their proposed optimal allocation procedures are not entirely dissimilar either, with the final allocations being highly dependent on whether the response rates for each collection mode are known or estimated using auxiliary data rather than the previously collected responses.

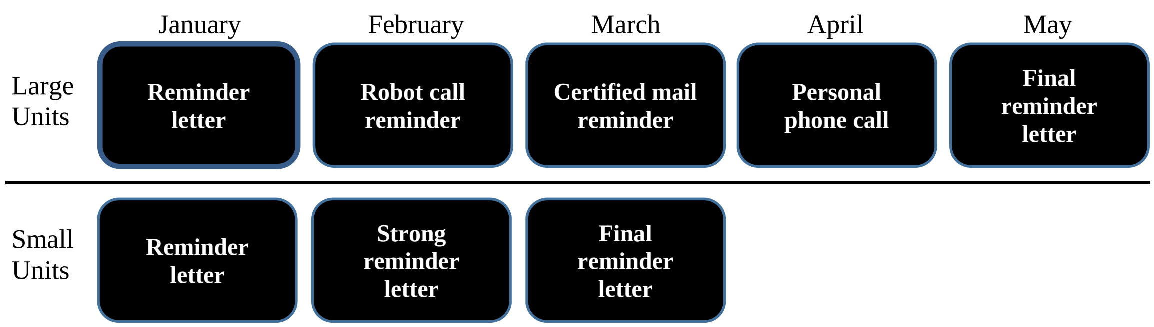

Figure 1.1 Nonresponse follow-up procedures for differing types of business in a fictional survey.

Fitting nonrespondent subsampling into a responsive or adaptive design framework is straightforward. As originally proposed by Groves and Herringa (2006), responsive designs require a minimum of two distinct phases of collection, with the second phase often being a probability subsample of nonrespondents that occurs at the "phase capacity" when the survey estimates are no longer changing, providing evidence the existing collection protocol is no longer cost-effective. Schouten, Calinescu and Luiten (2013) characterize responsive designs as a special case of adaptive collection designs. With an adaptive collection design, the data collection procedures can change (*adapt*) during the collection period. Paradata and sample data are used to determine whether to change the current procedures. The overall budget is fixed, but the implementation of a given strategy depends on (1) the realized sample of respondents at a point in time, (2) informative data obtained during data collection about the respondents and nonrespondents, and (3) information known in advance about the survey unit from the sampling frame. Consequently, selecting a probability sample of nonrespondents for NRFU – instead of attempting to contact all nonrespondents – falls under the adaptive design umbrella, with paradata (specifically response status) used to determine the sampling frame and frame data (e.g., the unit's size and industry classification) used as the basis of the sample design.

The U.S. Census Bureau is investigating nonrespondent subsampling strategies for the 2017 Economic Census (EC). Although a single program, the EC employs different sampling designs by sector (Probability proportional to size for the Construction sector, cut-off sampling for the Manufacturing and Mining sectors in collections prior to 2017, complete enumeration for the Wholesale Trade sector, and stratified simple random sampling without replacement (SRS-WOR) in the remaining sectors). Moreover, as is typical with many business programs, it is a multi-purpose collection, with the general statistics items collected from all surveyed units in a sector: examples include – but are not limited to – receipts/shipments, annual and first quarter payroll, and total employment. In addition, the EC collects information on product sales, types of which differ by sector and often industry. Imputation procedures differ by item, as do the estimators. Consequently, the subsampling design must be robust to sampling and estimator to the largest extent possible. We consider a *systematic sample* of nonrespondents sorted by a measure of size, a sampling design known to be as efficient as stratified simple random sampling without replacement (SRS-WOR) on average if the list is in random order and more efficient if the list is monotonic increasing or decreasing (Zhang, 2008; Lohr, 2010, Chapter 2, pages 50-51).

Ideally, the nonrespondent subsampling allocation procedure should be informed by properties of the respondent sample *during* the collection period. Of course, if the program is designed to collect one or two key items, then the allocation procedures should (at least attempt to) directly incorporate information on the survey design and estimation procedure, as well as detailed cost information, as proposed in Hansen and Hurwitz (1946) long-ago. In this case, one should use an optimal allocation procedure that minimizes costs subject to (estimable) reliability constraints. See Harter, Mach, Chaplin and Wolken (2007) and Beaumont, Bocci and Haziza (2014) for examples.

Such optimization is difficult to accomplish in the considered multi-purpose survey setting, especially when strongly correlated auxiliary variables are not available for all items. However, the OMB Statistical Standards for federal surveys require “survey (design) to achieve the highest practical rates of response, commensurate with the importance of survey uses, respondent burden, and data collection costs” and mandate nonresponse bias analyses for programs that fail to achieve these rates (Federal Register Notice, 2006). For nonrespondent subsampling occurring *during* the data collection cycle, imposing mandated lower bounds on the program-level response rate and in specified domains (examples include sampling strata or other post-strata such as industry code or type of government) is therefore a natural constraint to include in the allocation procedure.

In this paper, we explore allocation approaches that address such constraints, with an overall objective of selecting larger systematic subsamples in domains that have lower-than-targeted response rates. We introduce two optimized allocation procedures, both formulated as quadratic programs and solved with standard software packages: one that minimizes deviations between domain unit response rates and one that minimizes deviations between domain subsampling intervals. Our case study compares the statistical properties of subsamples obtained from each proposed allocation with three different estimators, considering two ratio estimators commonly used by business surveys along with the simple expansion (Horvitz-Thompson) estimator. The latter is not necessarily the most precise estimator when highly correlated auxiliary data are available, but gives an “upper bound” on the variance increase due to subsampling. The ratio estimators were selected to illustrate that the subsampling variance component can be reduced by incorporating correlated auxiliary data at the estimation stage.

Note that the presented allocation procedure is designed specifically for business surveys and implicitly assumes that largest units are excluded from the subsampling. In this case, the overall cost savings may not be substantial because the majority of a program's NRFU budget will be likely allocated to obtaining responses from the designated larger cases. However, the estimate quality can be improved. By equalizing response rates in considered domains, we hope to reduce the bias of the estimates by obtaining a respondent set that resembles the parent sample. Moreover, equalizing the subsampling intervals should help avoid overly increasing the sampling variance due to the second phase of selection, an unpleasant side effect of the additional stage of sampling that can completely offset any bias reduction obtained via the probability subsample (Biemer, 2010). And, it may be possible to further reduce both nonresponse bias and subsampling variance via an improved ratio or regression estimation procedure, if related covariates are available.

Section 2 provides context, briefly introduces the studied estimators, and presents our allocation procedures. Section 3 presents a simulation study that compares the statistical properties of the considered estimators for each realized allocation. We conclude in Section 4 with recommendations and suggestions for future research.

2 Methodology

2.1 Survey design and estimation

The general framework for our research is the two-phase sample design shown in Figure 2.1. The first stage is a stratified probability sample with a total sample size of n from a finite population (frame) of size N, performed *before* data collection begins. The survey is conducted, and units either respond or do not. During the data collection, response rates are monitored in H domains, where the domains do not necessarily equal the sampling strata. For example, total response rates might be monitored by three-digit industry classification, although these industry sampling strata are further broken down by size class. Furthermore, the domains could be independent of the original sampling strata e.g., race or sex categories (resembling post-strata). Hereafter, the term "domain" refers to the nonrespondent subsampling strata, indexed by $h\,(h = 1, 2, \ldots, H)$.

The second stage of probability sampling occurs at a predetermined point in the data collection cycle when we select an overall $1 - \text{in} - K$ subsample of size m_1 from the m nonrespondents (a two-phase sample); this predetermined point can be a fixed calendar date or via a responsive/adaptive design protocol. The value of K is determined by the program managers, who take into account the overall budget for NRFU (assumed fixed), mandated performance measures (e.g., response rates, coefficient of variation requirements), and other operational considerations such as length of collection period and available resources. Our allocation procedure determines the $1 - \text{in} - K_h$ systematic subsample of size m_{1h} from the m_h nonrespondents in each domain. Only the sampled m_{1h} units receive NRFU.

Our objective is to estimate Y, the population total of characteristic y. This estimate is $\hat{Y} = \hat{Y}_{R1} + \hat{Y}_{R2}$ where $\hat{Y}_{R1}$ is estimated from the r_{1h} first-stage sample respondents and $\hat{Y}_{R2}$ is estimated from the r_{2h} second-stage sample respondents (see Figure 2.1). Nonresponse adjustments to the r_{2h} subsampled (responding) units assume a missing at random response (MAR) mechanism, treated as a Bernoulli sample (Särndal, Swensson and Wretman, 1992, Chapter 15; Kott, 1994). We consider three different adjustment-to-sample reweighting estimators of $\hat{Y}_{R2}$ (Kalton and Flores-Cervantes, 2003): the double reweighted expansion (DE) estimator (Binder, Babyak, Brodeur, Hidiroglou and Wisner, 2000; Shao and Thompson, 2009; Haziza, Thompson and Yung, 2010), a separate ratio (SR) estimator that adjusts for unit nonresponse using a covariate that is highly correlated with both response propensity and the survey characteristic of interest (Shao and Thompson, 2009; Haziza et al., 2010), and a combined ratio (CR) estimator (Binder et al., 2000). Formulae are provided in the Appendix.

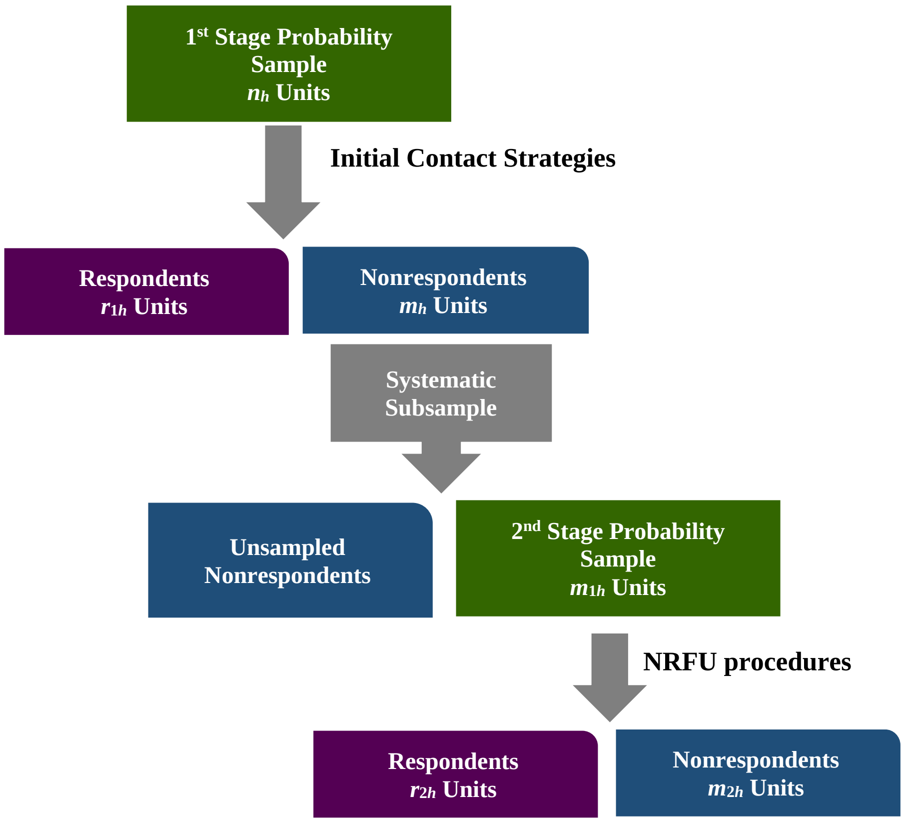

Figure 2.1 Nonrespondent subsample from probability sample, selected during data collection (two-phase sample design). Unsampled nonrespondents do not receive NRFU.

These estimators require a *minimum* of $r_{2h} = 1$ in each domain and a minimum of $r_{2h} = 2$ for variance estimation. These minimal conditions may not hold for several reasons. During the early stages of NRFU collection, an insufficient number of the subsampled units might respond in a given domain. Alternatively, the allocation procedure could determine that no subsampling is required in one or more domains. Lastly, the allocation procedure could require 100-percent follow-up (*all* units subsampled) in selected domains; henceforth, we refer to 100-percent follow-up/no subsampling as "full follow-up". In these cases, the estimation procedure ignores the last stage of sampling as if it did not occur and produces estimates for domain h using the collapsed estimator formulae provided in the Appendix.

2.2 Allocation strategies

When *all* nonresponding cases are subjected to NRFU, respondent contact strategies focus on improving overall response rates. Analysts might focus primarily on obtaining responses from soft refusal cases that they believe have similar characteristics to previous respondents ("quick wins"), although this phenomenon is more likely when the survey collection is performed in the field, as with household or agricultural surveys, and perhaps is less likely for internet or mail collections. With business surveys, the size of the unit is a factor in the NRFU procedures as discussed in Section 1.

Our objective is to obtain a realized set of respondents that approximates a random subsample of the originally selected sample via a probability sample of nonrespondents. With a probability sample, the targeted cases represent a cross-section of the nonrespondent population. By focusing contact efforts on the subsample, we hope to decrease the effects of nonresponse bias on the estimated totals by obtaining data from all types of nonresponding units. Moreover, weighting or imputation methods may be more effective at reducing the nonresponse bias effects with a probability subsample of nonrespondents (Brick, 2013). Even though they do not receive additional NRFU, the unsampled nonrespondent cases may provide responses later in the collection cycle. If so, an unbiased estimation procedure would not include the unsampled late responses in the final estimate assuming that all subsampled units respond, as these units are represented by the subsampled cases. However, this procedure is extremely distasteful to many survey managers. Instead, we include their data in the tabulations as if they had responded *before* subsampling. This does induce bias in the estimate. In practice, we ensure that this situation occurs infrequently by subsampling late in the data collection cycle.

With a business survey that keeps track of little or no demographic information, most of the information on the nonrespondents such as industry and unit size (e.g., total payroll, total receipts) is obtained from the sampling frame. Sorting the nonrespondents within prespecified domains by unit size and selecting a systematic sample should yield a subsample that resembles the originally designed sample in terms of unit size composition. This is especially important for business surveys where responses tend to be obtained from the larger units (Thompson and Washington, 2013). The choice of subsampling domain is determined by overall survey objectives such as publication levels or by the adjustment cell design (e.g., weighting cells or imputation classes), although computations are considerably simplified when the domain of interest is the original sampling strata. In the EC, the industry is the domain of interest.

We consider two allocation approaches: (1) equal-probability sampling; and (2) optimized allocation with constraints on unit response rates and sample size in predetermined domains. Equal probability sampling is easy to implement and should have the lowest sampling variance among considered nonrespondent subsampling allocation strategies, since the subsampling weight adjustment will be a constant value in all domains. However, since the same proportion of nonrespondents is sampled in each domain, the subsample may not be large enough to offset nonresponse bias effects on totals in low-responding domains. We refer to the allocations obtained by equal probability sampling as $\text{Constant}-K$, where K refers to the overall sampling interval $(1-\text{in}-K)$.

Our optimized allocation methods address the above concern by concentrating NRFU efforts in domains that have low response rates, attempting to select sufficient cases to achieve the performance benchmarks. This strategy may decrease the nonresponse bias in the totals if the response mechanism is MAR, conditional on the auxiliary variables used to define the domains; see Wagner (2012). However, it can increase the variance, as the subsampling intervals will differ and the weights will become more variable. To minimize the additional sampling variance caused by differing sampling intervals, the domain *nonrespondent* *subsampling* intervals should be close to the overall nonrespondent subsampling interval. To control costs, the allocation should not select more units for NRFU than budgeted. Recall that the federal survey environment requires that target response rates be achieved or nearly achieved, which makes all domains “equally” important from a data collection viewpoint.

To describe the allocation procedures, we introduce additional notation:

Unit response rate: $$\mathrm{URR} = \frac{\sum_h (r_{1h} + r_{2h})}{\sum_h n_h}$$

Target response rate: $$\mathrm{URR}^{\mathrm{T}} = \frac{\sum_h r_{1h} + (q_h m_h / K)}{\sum_h n_h}$$

Target domain response rate: $$\mathrm{URR}_h^{\mathrm{T}} = \frac{r_{1h} + (q_h m_h / K_h)}{n_h}$$

with r_{1h} units of the n_h originally sampled units responding *before* subsampling, leaving m_h units available for subsampling in each domain. The unit response rate (URR) is the *actual* proportion of responding sampled units (Thompson and Oliver, 2012) and *does not* include an adjustment for subsampling. The target response rate $(\mathrm{URR}^{\mathrm{T}})$ used for allocation is the expected maximum obtainable URR for a given overall subsampling rate $K,$ with q_h representing the conditional probability of ultimately responding to the census/survey in domain $h,$ given that the unit did not respond prior to subsampling. In the allocation procedure, q_h can be modeled from historical data if available or can be assumed constant for a new survey or for sensitivity analyses.

We formulate optimized allocation as a quadratic program and consider two different objective functions. The first quadratic program minimizes the squared deviation of the target response rate in each domain $\mathrm{URR}_h^{\mathrm{T}}$ from the overall target unit response rate $\mathrm{URR}^{\mathrm{T}},$ subject to linear constraints on the size of nonrespondent sample. This objective function is analogous to the numerator of the Pearson chi-square goodness-of-fit test.

The second quadratic program minimizes the squared deviation in domain sampling intervals from the *overall sampling interval* (K) subject to linear constraints on the unit response rates in each domain and on the number of sampled nonrespondents. Thus, although the optimization procedure allows the sampling intervals to vary by domain, the program tries to avoid potentially large increases in variance caused by the deliberately introduced “disproportionate sampling fractions” referred to in Kish (1992). We refer to the allocations obtained from these quadratic programs as $\mathrm{Min-URR}$ and $\mathrm{Min-}K$ respectively.

Both quadratic programs are primarily deterministic. However, recall that at the allocation stage, we must estimate the number of subsampled units that will eventually respond in each domain. Both quadratic programs use Constraints (1) through (3) in Table 2.1. Constraint (4) is included in the $\mathrm{Min-}K$ allocation to ensure that the optimization solution is not $K_h = K$ for all domain $h.$ There are two limiting scenarios (preconditions) that are addressed before the $\mathrm{Min-}K$ optimization. First, domains whose $\mathrm{URR}_h^{\mathrm{T}} \geq \mathrm{URR}^{\mathrm{T}}$ *before* subsampling must be removed from the optimization problem $(K_h = \infty).$ Second, if the estimated unit response rate cannot be possibly achieved in a given domain for an assumed $q_h,$ then all units in the domain are selected for NRFU $(K_h = 1)$. The $\text{Min}-K$ optimization is applied to the remaining domains, requiring that these subsampled domains have expected URRs that meet or exceed the target URRs.

Using sample data containing respondents and nonrespondents, along with different specified values for q_h, we use the SAS® PROC NLP (The data analysis for this paper was generated using SAS software. Copyright, SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.) to solve the quadratic programs (obtaining the set of K_h). The realized allocations are not integer values, and the real valued intervals (K_h) were input to SAS® PROC SURVEYSELECT to select stratified systematic subsamples of nonrespondents. As noted by one reviewer, this yields a solution that is randomly rounded but constrained at the overall required sample size, and there may be some impact on reliability due to rounding error. Such effects were not studied in this paper.

Table 2.1
Optimized allocation quadratic programs

		Min−URR	**Min−*K***	**Purpose**
Objective Function		$\min\sum_h (\text{URR}_h^{\text{T}} - \text{URR}^{\text{T}})^2$	$\min\sum_h (K_h - K)^2 = \min\sum_h \left(\frac{m_h}{m_{1h}} - K\right)^2$	
Constraints	(1)	$K \le \sum_h m_h \big/ \sum_h m_{1h}$		Selected sample size cannot exceed overall 1-in-K sample size
	(2)	$m_h \big/ m_{1h} \ge 1$		Domain subsample cannot exceed number of nonrespondents in the strata
	(3)	$m_{1h} \ge 0$		Non-negativity constraint
	(4)	Not Applicable	$\frac{r_{1h} + q_h m_h}{n_h} < \text{URR}^{\text{T}}$ $\quad K_h = 1$ $\frac{r_{1h}}{n_h} \ge \text{URR}^{\text{T}}$ $\quad K_h = \infty$ $\text{URR}_h^{\text{T}} \ge \text{URR}^{\text{T}}$ $\quad$ otherwise	Ensures that all domains achieve target URR as feasible.

3 Case study

This section presents the results of a simulation study that evaluates the considered allocation procedures on empirical sample data from the Annual Survey of Manufactures (ASM) from the 2010 and 2011 data collections. For more information on the ASM, see http://www.census.gov/manufacturing/asm.

The ASM is an establishment survey designed to produce "sample estimates of statistics for all manufacturing establishments with one or more paid employee(s)" (http://www.census.gov/manufacturing/asm/); it is a Pareto-PPS sample of approximately 50,000 establishments selected from a universe of 328,500. Approximately 20,000 establishments are included with certainty, and the remaining establishments are selected with probability proportional to a composite measure of size. Selected units are in the sample for the four years between censuses. Sampling strata are defined by six-digit industry code using the North American Industry Classification System.

The ASM estimates totals with a difference estimator (Särndal et al., 1992). To reduce respondent burden, units below a certain threshold are dropped from the sampling frame entirely. Instead, their data are imputed using administrative data values for selected items and industry-level regression models for the remaining items. Similarly, the ASM imputes complete records for unit nonrespondents. See http://www.census.gov/manufacturing/asm/ for additional information on the ASM methodology.

Because the items collected by the ASM questionnaire are a subset of the EC's manufacturing sector items, the ASM is often used to pretest new EC processing or data collection procedures. With the ASM and the EC, implementing a probability subsample of nonrespondents for NRFU represents a major procedural change. The ASM NRFU procedures are very similar to the EC procedures. Because a given company can comprise several establishments, the sets of multi-unit (MU) establishments corresponding to the company can be designated for phone follow-up as well as other company completeness checks. In contrast, the NRFU procedures for the single unit (SU) establishments – establishments with one location and parent company – differ. The largest SU establishments are included with certainty (sampled with probability = 1) and may receive a personal phone call in selected domains. The sampled SU establishments ("SU noncertainty establishments") receive some reminders, but are very unlikely to receive a personal phone call.

Our simulation study examines one of the fourteen key ASM items and employs the double expansion estimate and the two ratio estimators described in the Appendix, not the difference estimator used in ASM production estimates. Consequently, our results should not be extrapolated to the ASM.

3.1 Simulation study design

Our simulation study compares the statistical properties of total shipment estimates obtained from the three considered nonrespondent subsampling designs over repeated samples, using three different estimators. Our *sampling frame of nonrespondents* is derived from the fully imputed 2011 ASM sample and is limited to the SU noncertainty establishments so that the overall ASM publication reliability requirements are maintained. The ratio estimators employ the sample-based values of annual payroll as an auxiliary variable. This variable is highly correlated with total shipments, but is subject to imputation. Note that we use the *complete* ASM sample (all MU and SU establishments) for the allocations but present the relative bias and MSE results for the subsampled domains (SU noncertainty establishments) only.

For the SU noncertainty establishments, the first NRFU attempt – consisting of a reminder letter – is historically very effective, so nonrespondent subsample selection occurs before the second NRFU attempt. The second NRFU attempt is generally more expensive (historically a package re-mail, although reminder letters via certified mail will be used in future collections). Nonrespondent subsampling of SU noncertainty establishments occurs *after* the second contact attempt (i.e., after the first NRFU attempt).

To perform the simulation, we removed all MU establishments and SU certainty establishments from the ASM sample data to create a frame, and then independently repeated the following procedure 5,000 times for each allocation procedure:

1. Using the estimated response propensities provided in Table 3.1, randomly induce nonresponse into the sample using a MAR response mechanism.

2. Sort the induced nonrespondents by sampling weight.
3. Select a stratified systematic sample using the nonrespondent domain subsampling rates for a given allocation strategy.
4. Simulate unit response for each round of NRFU. Table 3.1 provides the conditional response propensities used for each distinct NRFU contact phase. These statistics use paradata from the 2010 and 2011 ASM collections (Fink and Lineback, 2013). Hereafter, we refer to these conditional probabilities as "nonrespondent conversion rates". If the unit responded, the mode of response is randomly assigned using historical frequencies provided by subject matter experts. After assigning response status/response mode to each unit, compute cumulative collection cost, URR, and estimates.
5. For each allocation, repeat Step 4 until either ten rounds of follow-up have been conducted or the total budget has been expended. If funds are exhausted within a round, then NRFU ceases. Given that the fixed budget assumes that $1/K$ of the original set of nonrespondents will receive NRFU, the budget can be exhausted under full follow-up. The total budget is never expended before ten rounds of NRFU with nonrespondent subsampling, as the cost-per-unit of mailing a reminder letter is quite low. Our choice of a maximum of ten rounds of NRFU in the simulation was subjective; the purpose was to demonstrate that subsampling would facilitate additional contact efforts at no additional cost.

Table 3.1
Nonrespondent conversion rates for noncertainty single unit establishments by NRFU contact round used for simulation

Domain	Initial Response Probability	Nonrespondent Conversion Rates for a given Round of Nonresponse Follow-up									
		1	2	3	4	5	6	7	8	9	10
1	0.31	0.27	0.15	0.17	0.24	0.12	0.06	0.03	0.03	0.03	0.03
2	0.44	0.32	0.24	0.15	0.36	0.18	0.09	0.05	0.05	0.05	0.05
3	0.39	0.28	0.24	0.18	0.11	0.06	0.03	0.02	0.02	0.02	0.02
4	0.35	0.36	0.17	0.19	0.18	0.09	0.05	0.02	0.02	0.02	0.02
5	0.25	0.19	0.13	0.10	0.17	0.09	0.04	0.02	0.02	0.02	0.02
6	0.27	0.13	0.29	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
7	0.44	0.34	0.23	0.20	0.25	0.13	0.06	0.03	0.03	0.03	0.03
8	0.38	0.45	0.12	0.33	0.25	0.13	0.06	0.03	0.03	0.03	0.03
9	0.39	0.30	0.23	0.13	0.25	0.13	0.06	0.03	0.03	0.03	0.03
10	0.75	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11	0.28	0.23	0.12	0.18	0.15	0.07	0.04	0.02	0.02	0.02	0.02
12	0.36	0.30	0.21	0.15	0.31	0.16	0.08	0.04	0.04	0.04	0.04
13	0.39	0.22	0.19	0.13	0.23	0.12	0.06	0.03	0.03	0.03	0.03
14	0.37	0.36	0.16	0.06	0.45	0.22	0.11	0.06	0.06	0.06	0.06
15	0.41	0.32	0.22	0.19	0.26	0.13	0.06	0.03	0.03	0.03	0.03
16	0.40	0.34	0.22	0.23	0.32	0.16	0.08	0.04	0.04	0.04	0.04
17	0.34	0.26	0.18	0.10	0.21	0.11	0.05	0.03	0.03	0.03	0.03
18	0.40	0.31	0.18	0.10	0.18	0.09	0.04	0.02	0.02	0.02	0.02
19	0.37	0.29	0.20	0.19	0.23	0.11	0.06	0.03	0.03	0.03	0.03
20	0.40	0.28	0.21	0.15	0.18	0.09	0.04	0.02	0.02	0.02	0.02
21	0.36	0.27	0.20	0.14	0.23	0.11	0.06	0.03	0.03	0.03	0.03

The nonrespondent conversion rates in the majority of domains follow the same pattern: a decaying response probability followed by a slight increase in the fourth round due to a longer collection period. Domain 10 does not follow this pattern; it contained only four units that all responded before subsampling began. After the 4th round of NRFU, the nonrespondent conversion rates are reduced by half until they achieve the minimum allowable value of 0.02. The pattern reflects the findings of Olson and Groves (2012) (Olson and Groves (2012) postulate that the response propensities change over the collection cycle, especially as data collection protocols are modified. With the ASM, the reminder letters become more stringent at each NRFU contact phase. Likewise, the authors demonstrate that response propensities decline over the collection phase when a stable data collection protocol is used, as reflected in nonrespondent conversion rates). Mail and phone response propensity estimates were provided by subject matter experts, as were approximate costs by mode and an overall budget figure.

To evaluate the statistical properties of each allocation method for each estimator, we computed the relative bias and the mean squared error. The relative bias (RBE) for each estimate of total shipments at NRFU phase t for a given sampling overall interval (K), allocation method a (Constant$-K$, Min$-K$, Min$-$URR), eventual response probability q, and estimator e (DE, SR, CR) is

$$\operatorname{RBE}(Y)_{Kaqt}^{e} = 100 * \left[\left(\frac{\sum_{s=1}^{5,000} \hat{Y}_{Kaqts}^{e}}{5,000} \bigg/ Y\right) - 1\right]$$

where $\hat{Y}_{Kaqts}^{e}$ is the estimated total and Y is the population total shipments value.

The mean squared error at NRFU phase t for a given sampling interval, allocation method and estimator is

$$\operatorname{MSE}(Y)_{Kaqt}^{e} = \left[\sum_{s=1}^{5,000} \left(\hat{Y}_{Kaqts}^{e} - Y\right)^2\right] \bigg/ 5,000.$$

Since our simulation induces MAR response, the DE estimates should be approximately unbiased over repeated samples, whereas the two ratio estimates should not be. However, the DE estimates are expected to have large variance; using ratio estimators with a positively correlated auxiliary variable is expected to reduce this variance (i.e., increase the precision). Thus, examining the MSE provides insight into the bias-variance tradeoff.

3.2 Allocation

The simulation study uses data from the 2011 ASM collection. Input parameters for allocation were estimated from 2010 ASM collection data. Recall that the target URR applies to the entire ASM program and is not restricted to the subsampling domains - in our case, SU noncertainty establishments. Consequently, the certainty SU and MU unit counts obtained from the 2010 ASM data are included in the allocation programs in the r_{1h} as constants; the remainder of the r_{1h} represents the estimated count of responding SU noncertainty establishments after the first round of NRFU is completed. To ensure that each nonrespondent sampling domain contained sufficient numbers of units to obtain a feasible solution, we used three-digit industry as NRFU sampling domain instead of the six-digit industry used for the ASM sample design [Note: the determination subsampling domain was determined collaborative with the ASM program managers and methodologists].

Both quadratic programs require an estimated probability of eventually responding to follow-up (q_h) to compute the $\mathrm{URR}^{\mathrm{T}}$ (overall and by domain). To assess the sensitivity of the allocation procedure, we tested ten different constant values $(q_h = 0.10, 0.20, \ldots, 1)$, keeping the value constant across all domains. A similar approach can be taken when historic paradata are not available. In addition, we estimate the q_h directly from the 2010 ASM data. These estimates vary by 20-percent at three-digit industry level. However, the median of these is nearly 50-percent. Consequently, the allocation obtained using the estimated (historic-data) q_h values are very similar to those obtained with $q = 0.50$.

Approximately \$21,000 was allotted for NRFU of SU noncertainty establishments after subsampling. With full follow-up, the expected final unit response rate was approximately 79%. Using data from the 2007 EC, Bechtel and Thompson (2013) found that the target industry unit response rates of 70% could only be achieved in a $1 - \text{in} - 3$ subsample if the average unit response rate in the majority of EC industries was 60% or larger *before* follow-up begins. With the ASM, the response rate prior to subsampling was approximately 57%. Instead, we select an overall $1 - \text{in} - 2$ subsample, which would save approximately 50-percent of the allotted budget after five completed rounds of NRFU at the cost of a decrease expected response rate (69%). The additional five rounds of NRFU added approximately \$4,000 to the total cost without commensurate increases in response rate (70%). A larger subsample would be preferable in terms of quality, but is not cost effective.

For allocation, we obtain the $\mathrm{URR}^{\mathrm{T}}$, allowing the q_h to vary by domain. The maximum URR is always achieved with the $\mathrm{Min}-\mathrm{URR}$ quadratic program. Table 3.2 presents the target URRs and the allocation subsampling rates obtained from the $\mathrm{Min}-\mathrm{URR}$ quadratic program. A dash (-) indicates no subsample is selected for NRFU (a sampling interval of ∞). If $K = 1$, all units in the domain are selected for NRFU (full follow-up). A label of $q =$ <value> indicates that the eventual probability of respondent is the same constant value in all domains; values estimated from historical data are labeled as $q_h =$ Est. Recall that $\mathrm{URR}^{\mathrm{T}}$ includes all respondent units in the ASM sample, not just the noncertainty single units that are eligible for subsampling. Consequently, selected domains have achieved their target URRs *before* subsampling and are not considered as subsampling candidates in the allocation programs.

As the probability of eventually responding increases, this allocation tends to select smaller subsamples in increasing numbers of domains. When the probability of an eventual response (q_h) is small (20-percent or less), then the allocations sensibly tend towards no subsampling or full follow-up, focusing on obtaining sample from the few domains with the poorest response rates. As the probability of an eventual response increases, the amount of subsampling tends to increase as well. At 70-percent, almost half of the domains are allocated at least one sampled unit, thus spreading the allocated sample across several domains instead of concentrating in a few domains that have exceptionally poor response rates. Note that rates below 20-percent are (hopefully) unrealistic as are rates greater than 70-percent. Domain 10 has highly variable sampling rates regardless; because all four units responded before subsampling, the quadratic program selected *any* sampling rate because, in effect, it always subsamples zero cases.

Table 3.2
Min−URR Allocations (Sampling Intervals) (Program Level $K = 2$)

	Min−URR										
Domain	***q* = 10**	***q* = 20**	***q* = 30**	***q* = 40**	***q* = 50**	***q* = 60**	***q* = 70**	***q* = 80**	***q* = 90**	***q* = 100**	**q_h = Est**
1	-	-	-	-	-	-	-	81.63	9.23	5.40	-
2	-	-	-	-	-	-	3.88	2.26	1.71	1.44	-
3	-	-	9.32	3.40	2.58	2.19	1.98	1.86	1.77	1.71	2.12
4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.06	1.13	1.00
5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	-
7	-	-	-	-	-	-	-	14.95	9.26	7.10	-
8	-	-	-	-	-	-	-	-	-	-	-
9	1.00	1.00	1.22	1.44	1.61	1.76	1.89	2.01	2.12	2.22	1.62
10	1.03	30.26	30.37	30.26	30.46	29.90	30.51	29.04	1.00	10.03	10.04
11	-	-	-	-	-	-	-	-	-	-	-
12	-	-	-	-	-	-	-	-	-	-	-
13	-	-	-	-	5.00	2.94	2.29	2.01	1.88	1.78	2.91
14	-	-	-	-	-	-	-	-	-	-	-
15	7.86	4.45	3.22	2.57	2.42	2.32	2.28	2.30	2.35	2.40	2.38
16	1.00	1.35	1.46	1.46	1.49	1.51	1.53	1.57	1.62	1.66	1.66
17	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
18	-	-	-	-	-	-	-	-	-	37.95	-
19	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
20	1.00	1.00	1.00	1.16	1.34	1.49	1.63	1.75	1.87	1.97	1.35
21	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
URR^T	72.5%	72.9%	73.3%	73.7%	74.1%	74.4%	74.8%	75.2%	75.6%	76.0%	74.3%

Unlike the Min−URR quadratic program, the $\text{Min}-K$ quadratic program did not always obtain a solution for a given target URR because of the domain-level constraints on the target URRs. When this occurred, we incrementally lowered the target response rate until a feasible solution could be obtained. Table 3.3 presents the target URRs and the allocations obtained from the $\text{Min}-K$ quadratic program.

Both the allocation methods tend to designate the same domains for either no subsampling or for full follow-up. However, the two methods produce very different allocations for the *same* q_h in the *subsampled* domains. The $\text{Min}-K$ allocations avoid subsampling in domains that have already achieved their maximum estimated target URR, regardless of the probability of eventually obtaining a response, with 40- to 50-percent of the domains not being subsampled when $0.30 \le q_h \le 0.50.$ Otherwise, the subsampling tends to be split between full follow-up (all units selected) or subsampling at an approximately $1-\text{in}-2$ sampling rate. In short, the Min−URR allocations yield domain subsampling intervals that can differ considerably from the overall interval, as the allocation seeks to equalize the target URR in each domain. The resultant variability in sampling intervals can lead to large increases in sampling variance. Because the $\text{Min}-K$ objective function seeks to equalize sampling intervals, the domain subsampling intervals tend to be less variable and are generally close to the overall sampling interval.

Table 3.3
Min$-K$ Allocations (Sampling Intervals) (Program Level $K = 2$)

					Min$-K$ (Target $K = 2$)						
Domain	***q*** **= 10**	***q*** **= 20**	***q*** **= 30**	***q*** **= 40**	***q*** **= 50**	***q*** **= 60**	***q*** **= 70**	***q*** **= 80**	***q*** **= 90**	***q*** **= 100**	***q*** **= Est**
1	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	-	2.00	2.00	-
3	-	-	-	-	1.99	2.00	2.00	2.00	2.00	2.01	1.99
4	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.10	1.18	1.26	1.00
5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
7	-	-	-	-	-	-	-	-	-	2.06	-
8	-	-	-	-	-	-	-	-	-	-	-
9	1.00	1.32	1.44	1.72	1.90	1.99	1.97	1.96	1.96	2.09	1.90
10	-	-	-	-	-	-	-	-	-	-	-
11	-	-	-	-	-	-	-	-	-	-	-
12	-	-	-	-	-	-	-	-	-	-	-
13	-	-	-	-	-	1.99	1.99	1.98	1.98	2.04	-
14	-	-	-	-	-	-	-	-	-	-	-
15	2.52	2.23	2.36	1.90	1.76	1.97	1.92	1.90	1.90	2.27	1.76
16	2.17	2.08	1.71	1.83	1.90	1.97	1.97	1.96	1.96	2.09	1.90
17	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
18	-	-	-	-	-	-	-	-	-	-	-
19	-	2.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
20	1.00	1.00	1.11	1.36	1.57	1.75	1.90	1.97	1.97	2.06	1.59
21	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
URR ^{T}	71.0%	71.4%	72.3%	72.7%	73.1%	73.4%	73.8%	74.2%	74.6%	75.0%	73.3%

3.3 Results

Our baseline closely mimics the NRFU procedures used in the 2012 ASM NRFU – four phases of full follow-up $(K = 1) -$ but can include an additional incomplete fifth round when the planned budget was not depleted to retain programming consistency. For other values of K, NRFU is concluded after ten rounds regardless of the remaining funds.

Table 3.4 presents the relative bias of the estimates (RBE) and the mean squared error (MSE) results obtained with full NRFU and the Constant$-K$ allocation for each considered estimator. In all cases, the unbiased double expansion (DE) estimator yields unbiased estimates, whereas the ratio estimators are slightly biased as expected. With subsampling, the relative bias of the ratio estimators increases, whereas the DE estimator remains unbiased. Regardless of estimator, the additional stage of subsampling increases the sampling variance and consequently the MSE; the bias tends to remain unaffected because the subsampled units are a representative subsample at each round of follow-up.

With equal probability subsampling (Constant$-K$), a subsample may contain a few sampled cases in one or more domains. Although the subsampling weighting adjustment is not variable, the nonresponse adjustment factors can be quite large. The optimal allocations are designed to equalize response rates across domains, which can lead to occasionally "oversampling" in low-responding domains. Table 3.5 presents the RBE and the MSE for the Min$-$URR optimal allocations, using three different constant values of

$q\,(q = 0.30, 0.50, 0.70)$ and the domain specific rates estimated from historical data $(q_h =$ Estimated). In all scenarios, the DE estimates are unbiased, the CR estimates are slightly biased, and the SR estimates are the most biased. This repeats the RBE pattern shown in the $\text{Constant}-K$ allocation results. Moreover, the RBE estimates do not appear to be overly sensitive to values of q_h used in allocation. Again, even with the additional rounds of NRFU, the bias of the subsamples' estimates is larger than that obtained with full follow-up of nonrespondents. In all cases, the MSE of the estimates obtained from the optimal allocations are smaller than those obtained with the $\text{Constant}-K$ allocations.

Regardless of estimator, the bias decreases when eventually probability of responding is low. This seems a bit counterintuitive but is in fact a direct consequence of the subsampling allocation procedure. When the probability of obtaining an eventual response is low, the $\text{Min}-\text{URR}$ allocation tends to subsample all or no units in a domain. With full follow-up, all responding units within the same domain have the same nonresponse adjustment. With a subsample, *only* the responding subsampled units' weights are adjusted for nonresponse and subsampling, in turn occasionally creating extremely variable weights within domain. As the probability of an eventual response increases, then the optimal allocation has sample in more domains, and finer adjustments are possible. With that said, the CR estimators tend to produce the lowest MSEs, regardless of allocation.

Table 3.4
Summary of relative bias in percent of the estimate and MSE for $\text{Constant}-K$ allocations in $x10^{12}$

$\text{Constant}-K$ Relative Bias of the Estimate						
Percent	**K = 1 (Full)**			**K = 2**		
Contact	**DE**	**CR**	**SR**	**DE**	**CR**	**SR**
2	0.01%	0.03%	0.10%	0.00%	0.51%	1.43%
3	0.00%	0.03%	0.08%	-0.01%	0.29%	0.77%
4	0.00%	0.01%	0.06%	-0.02%	0.14%	0.40%
5	0.01%	0.02%	0.04%	-0.01%	0.12%	0.32%
6				-0.01%	0.11%	0.29%
7				0.00%	0.11%	0.28%
8				0.00%	0.11%	0.27%
9				0.00%	0.10%	0.25%
10				0.00%	0.10%	0.25%
$\text{Constant}-K$ Mean Squared Error						
x10^12	**K = 1 (Full)**			**K = 2**		
Contact	**DE**	**CR**	**SR**	**DE**	**CR**	**SR**
2	4.96	2.60	5.56	37.53	26.34	70.49
3	3.67	1.96	4.17	19.82	13.80	28.88
4	2.55	1.39	3.03	11.75	8.30	14.87
5	2.48	1.39	2.87	9.94	7.10	12.12
6				9.36	6.75	11.16
7				9.09	6.63	10.63
8				8.80	6.48	10.23
9				8.51	6.32	9.95
10				8.27	6.19	9.74

Table 3.5
Summary of relative bias of the estimate and MSE for Min−URR optimal allocations

Min URR RBE (Target $K = 2$)												
Percent	***q* = 0.30**			***q* = 0.50**			***q* = 0.70**			***q* = Estimated**		
Contact	**DE**	**CR**	**SR**	**DE**	**CR**	**SR**	**DE**	**CR**	**SR**	**DE**	**CR**	**SR**
2	-0.01%	0.06%	0.20%	0.01%	0.07%	0.36%	0.01%	0.08%	0.31%	-0.01%	0.08%	0.32%
3	0.00%	0.05%	0.16%	0.01%	0.05%	0.26%	0.01%	0.07%	0.23%	0.01%	0.07%	0.23%
4	0.00%	0.05%	0.14%	0.01%	0.05%	0.18%	0.01%	0.06%	0.19%	0.01%	0.06%	0.17%
5	0.00%	0.05%	0.14%	0.01%	0.05%	0.18%	0.01%	0.05%	0.17%	0.00%	0.06%	0.15%
6	0.01%	0.05%	0.13%	0.02%	0.05%	0.17%	0.01%	0.05%	0.17%	0.00%	0.06%	0.15%
7	0.01%	0.05%	0.13%	0.01%	0.05%	0.17%	0.01%	0.05%	0.16%	0.00%	0.06%	0.15%
8	0.01%	0.05%	0.13%	0.01%	0.05%	0.16%	0.01%	0.05%	0.16%	0.00%	0.05%	0.14%
9	0.01%	0.05%	0.13%	0.01%	0.05%	0.16%	0.01%	0.05%	0.16%	0.00%	0.05%	0.14%
10	0.00%	0.05%	0.13%	0.01%	0.05%	0.15%	0.01%	0.05%	0.15%	0.00%	0.05%	0.14%
Min URR MSE (Target $K = 2$)												
x10^12	***q* = 0.30**			***q* = 0.50**			***q* = 0.70**			***q* = Estimated**		
Contact	**DE**	**CR**	**SR**	**DE**	**CR**	**SR**	**DE**	**CR**	**SR**	**DE**	**CR**	**SR**
2	12.55	6.77	16.03	14.35	7.48	17.60	14.31	7.44	17.15	14.39	7.79	17.05
3	8.88	5.13	10.87	9.80	5.43	11.32	9.57	5.42	11.36	9.75	5.47	10.98
4	7.00	4.28	8.15	7.61	4.45	8.28	7.31	4.43	8.44	7.53	4.44	8.05
5	6.61	4.07	7.41	7.02	4.17	7.44	6.80	4.13	7.60	6.94	4.14	7.42
6	6.45	3.97	7.16	6.78	4.09	7.18	6.62	4.03	7.25	6.75	4.05	7.15
7	6.37	3.92	7.05	6.68	4.06	7.08	6.55	3.97	7.07	6.67	4.02	7.03
8	6.34	3.90	6.94	6.57	4.01	6.97	6.45	3.93	6.95	6.57	3.98	6.93
9	6.28	3.87	6.86	6.50	3.98	6.89	6.39	3.90	6.86	6.47	3.94	6.84
10	6.23	3.85	6.78	6.40	3.91	6.76	6.35	3.87	6.75	6.42	3.89	6.73

The $\text{Min}-K$ allocation procedure is designed to reduce the variability in the subsampled units' adjustment weights. Table 3.6 presents the relative bias of the estimate and MSE for the $\text{Min}-K$ optimal allocation method. The $\text{Min}-K$ estimators display the same pattern as before. The DE estimates are unbiased, the CR estimates are nearly unbiased and the SR estimates are slightly biased.

The MSE estimates for the $\text{Min}-K$ method follow a similar pattern as the $\text{Min}-\text{URR}$ method, as expected due to the similarities between corresponding $\text{Min}-\text{URR}$ and $\text{Min}-K$ allocations. These results appear to be relatively insensitive to assumed eventual probability of response (q). The historical-data estimated conversion rates produce similar results to an assumed $q = 0.50$. In many cases, the $\text{Min}-\text{URR}$ method produces the least biased estimates. However, bias is only a single component of the MSE, and the $\text{Min}-\text{URR}$ allocations tend to have smaller expected number of respondents in several strata than their $\text{Min}-K$ counterparts. Moreover, the $\text{Min}-K$ allocations have smaller sampling variances by design, ultimately yielding estimates with lower MSEs than their $\text{Min}-\text{URR}$ counterparts.

Figures 3.1 and 3.2 plot the RBEs and MSEs obtained at each round of NRFU for the CR estimator (our "best" estimator) using the q_h obtained from historical data for each of the considered optimal allocation methods along with the benchmark values (labeled as "Full Follow-up"). In Figure 3.1, the benchmark estimates are the least biased. However, this extremely low bias is in part a consequence of our nonresponse model, which is uniform within domain and NRFU phase. Neither of the optimal allocation estimates attained the benchmark estimate levels, but they become very close after seven rounds of NRFU and the

RBEs of the Min−URR and Min−K CR estimates are less than *one tenth* of one percent (0.06% and 0.05% respectively). In summary, subsampling with either optimal allocation strategy yielded trivial biases increases over full follow-up.

Table 3.6
Summary of relative bias of the estimate and MSE for Min−K optimal allocations

Min−K RBE (Target K = 2)												
Percent	***q* = 0.30**			***q* = 0.50**			***q* = 0.70**			***q* = Estimated**		
Contact	**DE**	**CR**	**SR**	**DE**	**CR**	**SR**	**DE**	**CR**	**SR**	**DE**	**CR**	**SR**
2	0.03%	0.08%	0.24%	0.03%	0.09%	0.31%	0.00%	0.08%	0.33%	0.01%	0.07%	0.30%
3	0.03%	0.05%	0.20%	0.03%	0.08%	0.22%	0.00%	0.05%	0.22%	0.01%	0.06%	0.21%
4	0.02%	0.04%	0.16%	0.03%	0.07%	0.18%	0.00%	0.05%	0.17%	0.01%	0.05%	0.17%
5	0.02%	0.04%	0.15%	0.03%	0.06%	0.17%	0.01%	0.05%	0.16%	0.01%	0.05%	0.16%
6	0.02%	0.05%	0.14%	0.02%	0.06%	0.16%	0.01%	0.05%	0.15%	0.00%	0.05%	0.15%
7	0.02%	0.05%	0.14%	0.02%	0.05%	0.16%	0.01%	0.05%	0.15%	0.01%	0.05%	0.15%
8	0.02%	0.05%	0.14%	0.02%	0.05%	0.16%	0.01%	0.05%	0.15%	0.01%	0.04%	0.14%
9	0.02%	0.05%	0.14%	0.02%	0.05%	0.15%	0.01%	0.05%	0.14%	0.01%	0.04%	0.14%
10	0.02%	0.05%	0.14%	0.02%	0.05%	0.16%	0.01%	0.05%	0.15%	0.01%	0.04%	0.14%
Min−K MSE (Target K = 2)												
x10^12	***q* = 0.30**			***q* = 0.50**			***q* = 0.70**			***q* = Estimated**		
Contact	**DE**	**CR**	**SR**	**DE**	**CR**	**SR**	**DE**	**CR**	**SR**	**DE**	**CR**	**SR**
2	12.86	7.19	15.85	13.81	7.42	16.80	15.09	8.34	18.00	13.43	7.19	16.07
3	8.74	5.04	10.26	9.32	5.38	10.82	10.45	5.89	11.38	9.25	5.30	10.69
4	6.92	4.07	7.65	7.26	4.26	7.92	7.84	4.60	8.19	7.22	4.33	7.93
5	6.50	3.85	7.07	6.77	4.05	7.33	7.23	4.28	7.47	6.65	4.06	7.21
6	6.32	3.80	6.80	6.57	3.94	7.02	7.02	4.19	7.28	6.45	3.95	6.91
7	6.23	3.76	6.69	6.49	3.88	6.91	6.90	4.15	7.16	6.31	3.91	6.78
8	6.21	3.73	6.61	6.39	3.84	6.82	6.78	4.10	7.06	6.23	3.87	6.68
9	6.16	3.70	6.54	6.35	3.79	6.71	6.68	4.05	6.93	6.15	3.83	6.57
10	6.10	3.66	6.43	6.24	3.74	6.62	6.60	3.98	6.87	6.11	3.80	6.48

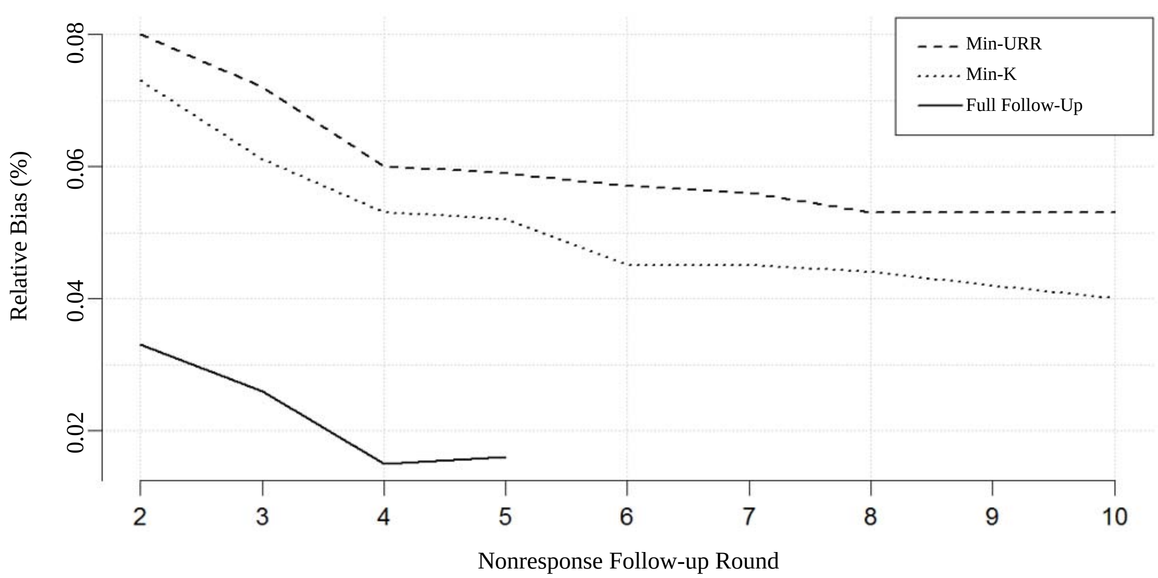

Figure 3.1 Relative bias of the estimates (Historic q_h) for the CR estimator.

Figure 3.2 plots MSE values by NRFU round using the CR estimator. The targeted nonresponse sampling strategy used for the $\text{Min}-K$ allocation appears to reduce the overall error. We believe that this is due to two factors. First, the $\text{Min}-K$ allocation procedure samples larger proportions of nonrespondents in low responding areas than obtained with the $\text{Min}-\text{URR}$ allocations. Second, the quadratic formula for the $\text{Min}-K$ allocation includes a constraint on the domain response rates, lowering the overall target response but reducing the variability in the proportion of respondents by domain. Ultimately, this approach yields similar response rates across sampling domains, indicative of a representative sample (Wagner, 2012; Schouten, Cobben and Bethlehem, 2009). Note that the increased MSE is not trivial with nonrespondent subsampling, even when using an adjustment procedure that benefits from a strong covariate in the ratio adjustment procedure. This is an acknowledged price paid for nonrespondent subsampling (Biemer, 2010). However, this additional variance component is measurable. If the measured component is too large, the program managers can subsample less (use a larger K).

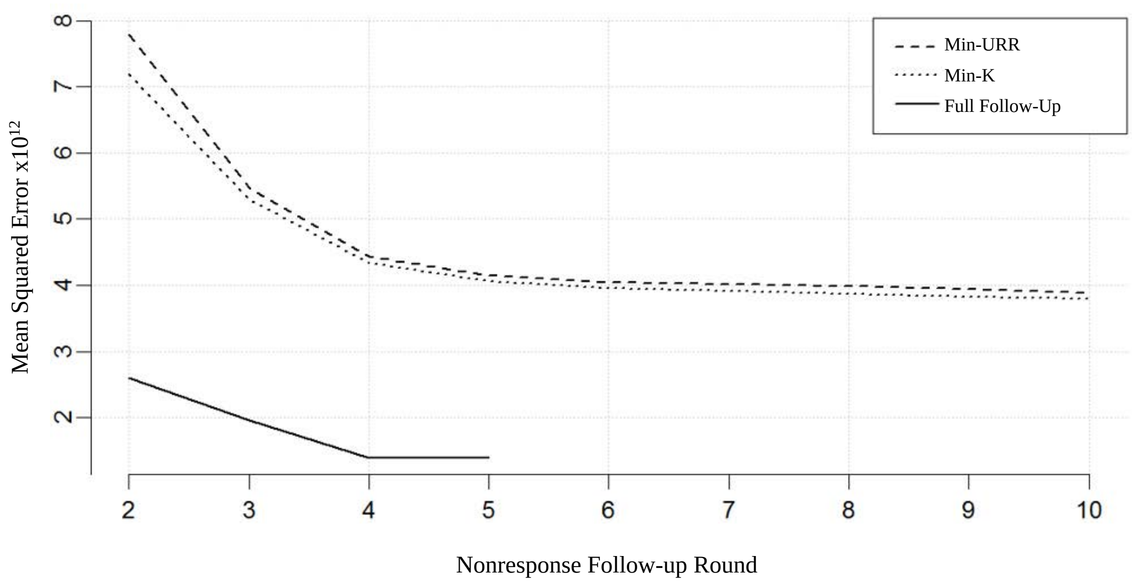

Figure 3.2 Mean squared error (Historic q_h) for the CR estimator.

3.4 Discussion

Given a sophisticated allocation method, a ratio estimator employing a highly correlated auxiliary variable, and a fairly large subsample, this case study shows that nonrespondent subsampling does not overly penalize quality to save cost. The additional stage of sampling increased the MSE for the studied variable, but the level was reduced by the judicious choice of estimator. Of course, we consider only one variable in our simulation, and this variable may or may not "behave" similarly to other survey items. One referee suggested the usage of an R-indicator (Schouten et al., 2009) or balance indicator (Särndal and Lundquist, 2014) to assess the overall representativeness of the respondent sets in a field survey setting. This might be useful at later stages of data collection (after nonrespondent subsampling and during NRFU), but would not provide any further insight into the degree of bias reduction on any collected item, as we can do in this simulation setting.

Of the three considered allocation methods, the $\text{Constant}-K$ method had the worst performance, often selecting a very small probability subsample when not needed and consequently increasing the sampling variance without reducing the bias. Of the three considered allocation methods, the $\text{Min}-K$ allocation was the most effective in realizing acceptable response rates and achieving reliable estimates; the larger bias caused by the varying domain sampling intervals is generally offset by the reduced sampling variance. However, implementation of the $\text{Min}-K$ allocation can be more challenging than the $\text{Min}-\text{URR}$.

For both optimal allocation procedures, we tested four different eventual probabilities of response to assess the sensitivity of the allocation procedures to these inputs. By comparing allocations obtained with a constant assumed input value to those obtained using the empirical estimates, we found that the realized allocations could over- or under- sample in selected domain, and the domain response rates could vary more than expected when the actual (survey) values are quite different from the input values. Consequently, we recommend using values estimated from historic paradata whenever possible.

If reducing cost is the overall goal, then we note that additional NRFU contact attempts beyond the fifth contact did not improve the bias or MSE of the subsampled estimates in our case study. Of course, if the achieved cost reduction for a $1 - \text{in} - 2$ subsample with up to ten NRFU contact attempts is acceptable, the funds allocated to these final contact attempts might be better expended earlier in the collection cycles using other contact strategies.

4 Conclusion

In general, the NRFU procedures for economic programs conducted by the U.S. Census Bureau follow a calendar schedule. Budget is tied to the fiscal year, and contact strategies are budgeted accordingly. Since economic populations are highly skewed and the statistics of interest are totals, a large fraction of the NRFU budget is allocated to the larger units. The smaller units are believed to be homogeneous – at least in size. However, it is difficult to validate that belief in the absence of collected respondent data. Given that the NRFU procedures rely on obtaining response data from the larger units, the response rates from smaller units tend to be much lower. It is quite likely that the realized respondent set is neither “balanced…which means (the selected sample has) the same or almost the same characteristics as the whole population” for selected items (Särndal, 2011) nor “representative… with respect to the sample if the response propensities ρ_i are the same for all units in the population” (Schouten et al., 2009). The emphasis on obtaining responses from the larger units at the cost of the lower unit response in turn creates a bias in the estimates, as imputed or adjusted values for smaller units resemble the large unit values (Thompson and Washington, 2013).

By limiting the target domain for nonrespondent subsampling to the smaller units, we can reduce this unmeasurable bias. Our allocation method increases the potential of obtaining a balanced and representative sample by targeting the low responding areas that usually would not receive any special treatment. It can be implemented at any stage of the data collection process and with any sample design, making it quite flexible although not necessarily optimal for specific sample designs and estimators. It is a “safe” approach for a multi-purpose survey, presumably designed to obtain reliable estimates for a variety of items. Moreover, selecting a systematic subsample from a list sorted by a unit measure of size avoids incidence of additional nonresponse bias incurred by focusing NRFU efforts on high response propensity cases (Tourangeau et al., 2016; Beaumont et al., 2014). We acknowledge that the increased variability in design weights and reduction in response rates are less than desirable effects caused by subsampling. However, these effects can be lessened via the choice of estimator, as demonstrated by our improved results with a ratio estimator. More sophisticated calibration estimators or other collapsed estimators could likewise be considered at the estimation stage.

Without probability subsampling, the contention that the realized respondent set of small businesses remains a probability sample is debatable. Several discussions of the summary report of the AAPOR Task Force on non-probability sampling (Baker, Brick, Bates, Battaglia, Couper, Dever, Gile and Tourangeau, 2013) specifically question whether “a probability sample with less than full coverage and high nonresponse should still be considered a probability sample”. That question is certainly relevant in our studied context, where sampled smaller units truly “opt in” to respond. Selecting a probability subsample of nonrespondents and instructing survey analysts to limit NRFU contact to these cases may limit this phenomenon. In addition, with a probability subsample, one can use accepted quality measures such as sampling error or response rates for evaluation.

All of the results presented for our case study assume that the existing NRFU contact strategies are used with the subsampled designs. However, subsampling nonrespondents without changing the data collection procedure may have minimal tangible benefits besides cost reduction. The reverse is also true: for example, Kirgis and Lepkowski (2013) present improved response data results for targeted small domains obtained with probability samples and revised contact strategies.

Tourangeau et al. (2016) note that “it is not always clear how to intervene to obtain cases, particularly cases with low underlying propensities, to respond”. This is especially relevant in the business survey context. Business surveys can draw on a wealth of cognitive research on data collection strategies for large companies: see Paxson, Dillman and Tarnai, 1995; Tuttle, Morrison and Willimack, 2010; Willimack and Nichols, 2010; Snijkers, Haraldsen, Jones and Willimack, 2013. In contrast, the smaller businesses receive very little personal contact (if any) and there is limited cognitive research on preferable contact strategies to draw upon. That said, the literature suggests that there are differences in collected data quality between large and small businesses: see Thompson and Washington (2013), Willimack and Nichols (2010), Bavdaž (2010), Torres van Grinsven, Bolko and Bavdaž (2014), and Thompson, Oliver and Beck (2015). Additional cognitive research for small establishments combined with field tests could yield better contact strategies. Subsampling nonrespondents paired with a new contact strategy for these “hard to reach” establishments would create a truly adaptive approach for all units, not just the larger ones. To this point, in response to these presented analyses, the Census Bureau conducted an embedded field experiment to test alternative NRFU strategies for selected small units in the 2014 ASM (Thompson and Kaputa, 2017). The outcome of that study was a new NRFU protocol implemented in the 2015 ASM and a second embedded field experiment that paired our proposed nonrespondent subsampling design with the most effective follow-up procedures determined from the 2014 test (Kaputa, Thompson and Beck, 2017).

Acknowledgements

Any views expressed are those of the author(s) and not necessarily those of the U.S. Census Bureau. The authors thank Eric Fink, Xijian Liu, Jared Martin, Edward Watkins III, Hannah Thaw, the Associate Editor, and two referees for their review of an earlier version of the manuscript, David Haziza for his thoughtful discussion of the paper, and Barry Schouten for his useful suggestions on the optimization problems.

Appendix

Our objective is to estimate Y, population total of characteristic y, from the realized sample of respondents. Let

- S_{hi} = 1 if unit i in domain h was in original sample; 0 otherwise.
- θ_{hi} = the probability of sampling unit i in domain h into the original sample $(w_{hi} = 1/\theta_{hi})$.
- R_{hi} = 1 if unit i in domain h provided a response before subsampling time t (value for y); 0 otherwise.
- I_{hi} = 1 if unit i in domain h was selected for NRFU (i.e., was a subsampled nonrespondent); 0 otherwise.
- J_{hi} = 1 if unit i in domain h responds, given selected into nonrespondent subsample; 0 otherwise.
- f_{hi} = adjustment factor for nonrespondent subsampling and unit nonresponse after NRFU.
- y_{hi} = value of characteristic y for unit i in domain h, available only for respondents.
- x_{hi} = value of characteristic x for unit i in domain h, available for all sampled units considered for nonrespondent subsampling (i.e., the nonrespondent subsampling frame). Then $\hat{Y} = \sum_h \sum_i w_{hi} y_{hi} S_{hi} R_{hi} + \sum_h \sum_i w_{hi} f_{hi} y_{hi} S_{hi} (1 - R_{hi}) I_{hi} J_{hi} = \hat{Y}_{R1} + \hat{Y}_{R2}$.

We consider three different adjustment-to-sample reweighting estimators of $\hat{Y}_{R2}$:

Double Expansion (DE): $$\hat{Y}_{R2}^{\mathrm{DE}} = \sum_h \sum_{i \in h} w_{hi} K_h \left(\frac{m_{1h}}{r_{2h}}\right) y_{hi} S_{hi} (1 - R_{hi}) I_{hi} J_{hi}$$

Separate Ratio (SR): $$\hat{Y}_{R2}^{\mathrm{SR}} = \sum_h \sum_{i \in h} w_{hi} K_h \left(\frac{\sum_{i \in m_{1h}} x_{hi}}{\sum_{i \in r_{2h}} x_{hi}}\right) y_{hi} S_{hi} (1 - R_{hi}) I_{hi} J_{hi}$$

Combined Ratio (CR): $$\hat{Y}_{R2}^{\mathrm{CR}} = \sum_h \sum_{i \in h} w_{hi} K_h \left(\frac{m_{1h}}{r_{2h}}\right) \left(\frac{\sum_{i \in m_{1h}} w_{hi} K_h x_{hi}}{\sum_{i \in r_{2h}} w_{hi} K_h \left(\frac{m_{1h}}{r_{2h}}\right) x_{hi}}\right) y_{hi} S_{hi} (1 - R_{hi}) I_{hi} J_{hi}.$$

Note that the DE and CR estimators are variations of the recommended reweighting procedure described in Brick (2013) and are discussed in Binder et al. (2000) among others. The DE estimator is the InfoS estimator presented in Särndal and Lundström (2005), studied in Shao and Thompson (2009), among others;

the SR estimator is a variation of the InfoP estimator presented in Särndal and Lundström (2005), treating the realized sample as the "population". Sampling weights were not included in the SR so that the adjustment reduces to the DE adjustment when $x_{hi} \equiv 1 \forall i \in h$; note that this unweighted response rate adjustment is recommended in Little and Vartivarian (2005). The CR estimator is presented in Binder et al. (2000), and is also studied in Shao and Thompson (2009). In our case study, a better choice might have been the quasi-randomization estimator from Oh and Scheuren (1983), which incorporates sampling weights in the adjustment factor, thus reducing their variability.

Collapsed estimators are used in three scenarios: (1) All units in the domain receive NRFU (no subsampling); (2) No units in the domain receive NRFU because response rate targets have been achieved (no subsampling); and (3) A single subsampled unit responded to NRFU (subsampling). The collapsed estimators analogues are given as follows:

Collapsed DE: $$\hat{Y}_h^{\text{DE,C}} = \sum_{i \in h} w_{hi} \left(\frac{n_h}{r_{1h} + r_{2h}} \right) y_{hi} S_{hi} R_{hi}$$

Collapsed SR: $$\hat{Y}_h^{\text{SR,C}} = \sum_{i \in h} w_{hi} \left(\frac{\sum_{i \in n_h} x_{hi}}{\sum_{i \in r_{1h} + r_{2h}} x_{hi}} \right) y_{hi} S_{hi} (1 - R_{hi}) I_{hi} J_{hi}$$

Collapsed CR: $$\hat{Y}_h^{\text{CR,C}} = \sum_{i \in h} w_{hi} \left(\frac{n_h}{r_{1h} + r_{2h}} \right) \left(\frac{\sum_{i \in n_h} w_{hi} x_{hi}}{\sum_{i \in r_{1h} + r_{2h}} w_{hi} \left(\frac{n_h}{r_{1h} + r_{2h}} \right) x_{hi}} \right) y_{hi} S_{hi} R_{hi}.$$

References

Baker, R., Brick, J.M., Bates, N., Battaglia, M., Couper, M., Dever, J., Gile, K. and Tourangeau, R. (2013). Summary report of the AAPOR task force on non-probability sampling – Report and rejoinder. *Journal of Survey Statistics and Methodology*, 1, 90-137.

Bavdaž, M. (2010). The multidimensional integral business survey response model. *Survey Methodology*, 36, 1, 81- 93. Paper available at https://www150.statcan.gc.ca/n1/pub/12-001-x/2010001/article/11245-eng.pdf.

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.

Bechtel, L., and Thompson, K.J. (2013). Optimizing unit nonresponse adjustment procedures after subsampling nonrespondents in the Economic Census. *Proceedings of the Federal Committee on Statistical Methods Research Conference*, https://nces.ed.gov/FCSM/index.asp.

Biemer, P. (2010). Total survey error: Design, implementation, and evaluation. *The Public Opinion Quarterly*, 74(5), 817-848.

Binder, D., Babyak, C., Brodeur, M., Hidiroglou, M. and Wisner, J. (2000). Variance estimation for two-phase stratified sampling. *The Canadian Journal of Statistics*, 28, 751-764.

Brick, J.M. (2013). Unit nonresponse and weighting adjustments: A critical review. *Journal of Official Statistics*, 29, 329-353.

Federal Register Notice (2006). OMB Standards and Guidelines for Statistical Surveys, Washington, DC.

Fink, E., and Lineback, J.F. (2013). Using paradata to understand business survey reporting patterns. *Proceedings of the Federal Committee on Statistical Methods Research Conference*, https://nces.ed.gov/FCSM/index.asp.

Groves, R., and Herringa, S. (2006). Responsive design for household surveys: Tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society Series A*, 169(3), 439-57.

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.

Harter, R.M., Mach, T.L., Chaplin, J.F. and Wolken, J.D. (2007). Determining subsampling rates for nonrespondents. *Proceedings of the Third International Conference on Establishment Surveys*, American Statistical Association.

Haziza, D., Thompson, K.J. and Yung, W. (2010). The effect of nonresponse adjustments on variance estimation. *Survey Methodology*, 36, 1, 35-43. Paper available at https://www150.statcan.gc.ca/n1/pub/12-001-x/2010001/article/11246-eng.pdf.

Kalton, G., and Flores-Cervantes, I. (2003). Weighting methods. *Journal of Official Statistics*, 19, 2, 81-97.

Kaputa, S., Thompson, K.J. and Beck, J. (2017). An embedded experiment for targeted nonresponse follow-up in establishment surveys. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Kirgis, N., and Lepkowski, J. (2013). Design and management strategies for paradata-driven responsive design: Illustrations for the 2006-2010 National Survey of Family Growth. *Improving Surveys with Paradata*, (Ed., Frauke Kreuter). Hoboken, NJ: John Wiley & Sons, Inc.

Kish, L. (1992). Weighting for unequal P_i. *Journal of Official Statistics*, 8(2), 183-200.

Kott, P. (1994). A note on handling nonresponse in sample surveys. *Journal of the American Statistical Association*, 89, 693-696.

Little, R.J., and Vartivarian, S. (2005). Does weighting for nonresponse increase the variance of survey means? *Survey Methodology*, 31, 2, 161-168. Paper available at https://www150.statcan.gc.ca/n1/pub/12-001-x/2005002/article/9046-eng.pdf.

Lohr, S.L. (2010). *Sampling: Design and Analysis*. Boston: Brooks/Cole.

Oh, H.L., and Scheuren, F.J. (1983). Weighting adjustment of unit nonresponse. *Incomplete Data in Sample Surveys*. New York: Academic Press, 20, 143-184.

Olson, K., and Groves, R.M. (2012). An examination of within-person variation in response propensity over the data collection field period. *Journal of Official Statistics*, 28, 29-51.

Paxson, M.C., Dillman, D.A. and Tarnai, J. (1995). Improving response to business mail surveys. In *Business Survey Methods*, (Eds., B.G. Cox, D. Binder, B. Nanajamma Chinnappa, M. Colledge and P. Kott). New York: John Wiley & Sons, Inc.

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 Lundquist, P. (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.

Särndal, C.-E., and Lundström, S. (2005). *Estimation in Surveys with Nonresponse*. Hoboken, NJ: John Wiley & Sons, Inc.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer Verlag.

Schouten, B., Calinescu, M. and Luiten, A. (2013). Optimizing quality of response through adaptive survey designs. *Survey Methodology*, 39, 1, 29-58. Paper available at https://www150.statcan.gc.ca/n1/pub/12-001-x/2013001/article/11824-eng.pdf.

Schouten, B., Cobben, F. and Bethlehem, J. (2009). Indicators for the representativeness of survey response. *Survey Methodology*, 35, 1, 101-113. Paper available at https://www150.statcan.gc.ca/n1/pub/12-001-x/2009001/article/10887-eng.pdf.

Shao, J., and Thompson, K.J. (2009). Variance estimation in the presence of nonrespondents and certainty strata. *Survey Methodology*, 35, 2, 215-225. Paper available at https://www150.statcan.gc.ca/n1/pub/12-001-x/2009002/article/11043-eng.pdf.

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

Thompson, K.J., and Kaputa, S. (2017). Investigating adaptive nonresponse follow-up strategies for small businesses through embedded experiments. *Journal of Official Statistics*, 33(3), 1-23.

Thompson, K.J., and Oliver, B. (2012). Response rates in business surveys: Going beyond the usual performance measure. *Journal of Official Statistics*, 27, 221-237.

Thompson, K.J., Oliver, B. and Beck, J. (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.

Thompson, K.J., and Washington, K.T. (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.

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), 579-606.

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 A*, 180, 203-223.

Tuttle, A., Morrison, R. and Willimack, D. (2010). From start to pilot: A multi-method approach to the comprehensive redesign of an economic survey questionnaire. *Journal of Official Statistics*, 26, 87-103.

Wagner, J. (2012). Research synthesis: A comparison of alternative indicators for the risk of nonresponse bias. *Public Opinion Quarterly*, 76(3), 555-575.

Willimack, D., and Nichols, E. (2010). A hybrid response process model for business surveys. *Journal of Official Statistics*, 26, 3-24.

Zhang, L.C. (2008). On some common practices of systematic sampling. *Journal of Official Statistics*, 24, 557-569.