Determining Allocation Requirements For Subsampling Nonrespondents From the Annual Survey of Manufactures

Daniel R. Whitehead, Stephen J. Kaputa, and Katherine Jenny Thompson[1]

U.S. Census Bureau
4600 Silver Hill Road; Washington, DC 20233
Daniel.Whitehead@census.gov

Proceedings of the 2013 Federal Committee on Statistical Methodology (FCSM) Research Conference

Abstract

As survey costs increase while response rates decrease, many agencies have begun to consider adaptive strategies for data collection. One such strategy is to select a probability sample of nonresponding units for follow-up instead of attempting complete follow-up on all nonrespondents. This report presents the results of a simulation study conducted to develop allocation strategies for selecting a subsample of nonrespondents for follow-up in the Annual Survey of Manufactures, given a fixed total cost and reliability constraints. The simulation study accounts for the two-phase sampling, testing the sensitivity of the assumed response levels and response mechanism, by examining the increase in variance (or coefficient of variation) caused by subsampling while monitoring follow-up cost.

1. Introduction

Consider a typical business survey, where data collection begins by mailing a form or a letter containing an internet address to each survey unit. During the data collection period, the survey organization keeps track of the returns. Often, the organization administering the survey (the "survey organization") establishes and maintains relationships with a select set of businesses to ensure that their forms are completed correctly and in a timely manner; usually, these personal contacts are reserved for larger businesses or those with very complex organizational structures (Snijkers et al., 2013). However, all sampled units that do not respond to the survey will receive some form (or forms) of reminder (nonresponse follow-up).

In our setting, the program has an annual fixed budget, and nonresponse follow-up costs are components in the overall budget. Consequently, nonresponse follow-up has a fixed calendar schedule. As a very simple example, the collection calendar might be:

- January 1: Forms mailed
- March 1: Reminder letter sent to all nonresponding units
- June 1: Package re-mailed and reminder phone calls are made
- September 1: Reminder letter is sent to remaining nonrespondents

Obviously, this is not a responsive design (Groves and Heeringa, 2006). Instead, the funds for each round are determined in advance using estimates based on historic information from prior collections. Thus, a high proportion of the follow-up costs are already expended before follow-up begins: letters and packages are printed in advance, and bulk mail is contracted. Likewise, a portion of the budget is set aside for telephone follow-up, especially when centralized phone call centers are used.

As survey costs increase while response rates decrease, many agencies have begun to consider adaptive strategies for data collection. In today's federal budget environment, the term "fixed budget" is a misnomer. If the actual dollar amount available is less than budgeted, an adaptive nonresponse follow-up strategy can mitigate the

[1] This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau.

detrimental impact on quality that could occur if the follow-up procedures were abruptly curtailed. Indeed, if the adaptive design strategy includes probability sampling of nonresponding units for follow-up, survey costs can be reduced without greatly reducing quality. One potential quality benefit of focusing on a more targeted number of units for nonresponse follow-up is the possible reduction in mean squared error (MSE) that occurs if the reduction in nonresponse bias offsets the increase in variance due to subsampling.

Unfortunately, there are bias and variance trade-offs associated with subsampling. Probability subsampling of nonrespondents adds an extra stage to the original sample design. This in turn adds a variance component, even under 100% response from the subsample. When designing this subsample, it is important to take the additional variance component into consideration, by both maximizing the sampling rate (to avoid overly increasing the sampling variance) and by attempting to maintain consistent subsampling rates within each sampling stratum to minimize the variability of the weights. Note that probability subsampling adds a measurable *sampling* error component. It is difficult – if not impossible – to measure the nonsampling error due to unit nonresponse under any follow-up scenario, although some nonsampling errors can be reflected in variance estimates obtained with replication methods. Of course, selection for follow-up does not ensure that the unit will respond. A study that assesses the feasibility of subsampling nonrespondents should include this uncertainty in the evaluation and should test the sensitivity of the assumptions.

In this paper, we develop an allocation procedure for subsampling nonrespondents that tracks the cumulative cost (denoted running cost) and the sampling variance after each round of follow-up, assuming a systematic sample of nonrespondents. The procedure employs a simulation approach that allows for varying response probabilities under a missing-at-random (MAR) response mechanism to assess sensitivity. We apply this procedure to empirical data collected from the 2010 Annual Survey of Manufactures (ASM), a Pareto-probability proportional to size (PPS) sample of establishments. Data collection procedures for the subsampled nonrespondents are not considered here and is a subject for separate research.

In Section 2, we describe our allocation procedures. Section 3 provides general background on the ASM and provides context for this research. We summarize our results in Section 4 and conclude in Section 5 with some general comments.

2. Allocation Procedure

We propose selecting a systematic sample of nonrespondents from a list sorted by unit measure of size. Systematic sampling can yield estimates that are more precise over simple random sampling when the sample exhibits a monotone trend. In this case, variance estimates from the systematic sample may be smaller but inestimable when compared to simple random sampling variance estimates (Lohr, 2010). Since business survey analyst phone follow-up procedures begin with the largest and most complex units and may not encompass the smaller units due to time constraints, the systematic sample should be more "representative" than the analogue obtained from an attempted 100% follow-up sample.

The allocation objective is to determine the "optimal" subsampling interval (K) that balances the simultaneous objectives of reducing data collection costs while minimizing the sampling variance of the estimate. To determine the best allocation rate for a program, we use the following simulation procedure, repeating steps 1 – 5 independently 5000 times:

1. Randomly induce nonresponse in the complete dataset using a MAR mechanism, dividing the sample into two groups: respondents (R_{ORIG}) and nonrespondents (NR_{ORIG}).
2. Sort the nonrespondents by measure of size.
3. Select a systematic sample with a rate of ($1/_K$) (once per replicate)
4. Simulate nonresponse follow-up. In a given round of nonresponse follow-up *t,* each unit will respond by data collection mode *m* with probability $\pi_{m,t}$, where $\sum_m \pi_{m,t} < 1$. Each mode has a different collection cost. After assigning response status to each unit, compute cumulative collection cost, estimate value, and sampling variance.

5. Repeat Step 4 until either ten rounds of follow-up have been conducted or the total budget has been expended. If exhaustion of funds occurs within a round, then follow-up will cease in that particular round, without pursuing all remaining sampled nonrespondents.

The reweighting and estimation procedure, explained below, is repeated independently at each round of follow-up within replication, allowing us to compute cost and variance estimates through time.

We use the modified Horvitz-Thompson estimator $(\hat{Y}_{HT})$ to estimate the population total for characteristic *y* shown in (1). Our estimator is the sum of two separate estimators. The first component $(\hat{Y}_{HT,ORIG})$ is the weighted sum of the initial respondents (R_{ORIG}) using the survey design weights. The second component $(\hat{Y}_{HT,SUB})$ represents the reweighted sum of the subsampled units that responded to follow-up efforts. The nonresponse adjustment weighting procedure is repeated independently at each round of follow-up within replicate. We use an adjustment-to-sample nonresponse adjustment procedure that multiplies sampling weights by the inverse response rate of the subsample (Kalton and Flores-Cervantes, 2003) in the ASM application described in Section 3. Although this estimator is appropriate under MAR, it does increase the variance of the estimate. Using a ratio adjustment procedure or another form of calibration with positively correlated covariates could reduce the variance induced by the adjustment cell weighting. For an example of ratio estimation in this context, see Bechtel and Thompson (2013). Formally, at each round of follow-up in replicate *s*, we estimate

$$\hat{Y}_{HT,s,t} = \sum_{i \in U} \frac{y_i}{\pi_{i,A,t}} I_{A,s,t} I_{R,s,t} + \sum_{i \in U} \frac{y_i}{\pi_{i,A,t}} I_{A,s,t} \left(1 - I_{R,s,t}\right) K\, I_{K,s,t} \frac{n_{K,s,t}}{n_{F,s,t}} I_{F,s,t}$$

Suppressing the notation for replicate *s* and round *t*, we estimate

$$\hat{Y}_{HT} = \sum_{i \in U} \frac{y_i}{\pi_{i,A}} I_A I_R + \sum_{i \in U} \frac{y_i}{\pi_{i,A}} I_A \left(1 - I_R\right) K\, I_K \frac{n_K}{n_F} I_F$$

$$= \sum_{i \in U} w_i y_i\, I_A I_R + \sum_{i \in U} \left(w_i K \frac{n_K}{n_F}\right) y_i\, I_A \left(1 - I_R\right) I_K\, I_F$$

$$= \sum_{i \in U} w_i y_i\, I_A I_R + \sum_{i \in U} \widetilde{w_i}\, y_i\, I_A \left(1 - I_R\right) I_K\, I_F$$

$$= \hat{Y}_{HT,ORIG} + \hat{Y}_{HT,SUB} \qquad (1)$$

Where
U = the universe (population)
y_i = the value of the characteristic for unit *i*
$\pi_{i,A}$ = the probability of unit *i*'s inclusion in the initial sample
I_A = 1/0 indicator variable for sample inclusion (=1 for all units in simulation)
I_R = 1/0 indicator variable for response in initial sample (I_R = 1 indicates response)
I_K = 1/0 indicator variable for subsample inclusion
I_F = 1/0 indicator variable for response to cumulative set of follow-up rounds
n_K = number of nonresponding units sampled for follow-up
n_F = number of subsampled units that responded to follow-up
w_i = the original sampling weight
$\widetilde{w_i}$= the adjusted sampling weight $\widetilde{w_i} = \left(w_i K \frac{n_K}{n_F}\right)$

Under 100% follow-up (*K* = 1) and complete response (sampled nonrespondents), (1) reduces to the standard Horvitz-Thompson estimator (Lahiri, 2012). In each replicate *s*, after completing a follow-up round, we also estimate running cost (cumulative) and an estimate of the variance of $\hat{Y}_{HT}$. The running total cost after each follow-up round is

$$TC = \sum_{t=0}^{T} MO_t \, n_t + \sum_{t=0}^{T} MR_t \, n_{m,t} + \sum_{t=0}^{T} PR_t \, n_{p,t} \tag{2}$$

Where
TC = total cost
MO_t = the mailout cost during round t
n_t = the number of units mailed a form or letter during round t
MR_t = the cost of processing response by mail during round t
$n_{m,t}$ = the number of units responding by mail in round t
PR_t = the cost of processing response by phone during round t
$n_{p,t}$ = the number of units responding by phone in round t
The estimated variance has three components: the variance from the respondents in the original sample$\left(\hat{Y}_{HT,ORIG}\right)$, the variance due to the estimate derived from the nonresponse follow-up respondents$\left(\hat{Y}_{HT,SUB}\right)$, and the covariance between the two estimates. Because initial response to the survey is a random variable, there is negative covariance between our initial estimator and our follow-up estimator. Our estimator of the variance of $\hat{Y}_{HT}$ is given by

$$v\left(\hat{Y}_{HT}\right) = v\left(\hat{Y}_{HT,ORIG}\right) + v\left(\hat{Y}_{HT,SUB}\right) + 2\,cov\left(\hat{Y}_{HT,ORIG}, \hat{Y}_{HT,SUB}\right)$$

Although the ASM is a Pareto-PPS sample (see Section 3.1), the publication variance estimates are obtained with the Poisson sampling variance formula. Consequently, we use the Poisson sampling formula in our simulation study. However, before presenting the components of our variance estimator, we define the following terms:

$\pi_{i,A}$ = the probability of unit i's inclusion in the initial sample
n_A = the number of units selected in the initial sample
R_A = the set of respondents in the initial sample
$\pi_{i,R}$ = the conditional probability of unit i responding to the initial sample given their inclusion in the sample
n_R = the number of respondents in the initial sample
nr_{NR} = the number of nonrespondents in the initial sample
$\pi_{i,K}$ = the conditional probability of nonresponding unit i being subsampled given their initial nonresponse
$\left(Assume\ E\left(\frac{n_K}{nr_{NR}}\right) = \frac{1}{K}, so\ that\ \pi_{i,K} = \frac{n_K}{nr_{NR}} \doteq \frac{1}{K}\right)$
n_K = the size of the follow-up systematic sample $\left(Assume\ n_K = \frac{nr_{NR}}{K}\right)$
$\pi_{i,F}$ = the conditional probability of unit i's responding to the follow-up sample, given their selection
$\left(Assume\ \pi_{i,F} = \frac{n_F}{n_K}\right)$
n_F = the number of responding units from the follow-up sample
F_t = The set of respondents after round t of follow-up has been completed
C_i, C_j = groups of term treated as a constant during a particular aspect of the variance estimation

The variance estimate of a Horvitz-Thompson estimate from a Poisson sample is given by

$$v\left(\hat{Y}_{HT,ORIG}\right) = \sum_{i \in R_{ORIG}} \frac{\left(1 - \pi_{i,A}\,\pi_{i,R}\right)}{{\pi_{i,A}}^2} \; {y_i}^2$$

See Cochran (1977), Särndal et al. (1992), and Lahiri (2012).

The systematic subsample of nonrespondents is a two-phase sample. A model-assisted perspective on nonresponse considers the "selection" of responding units from the nonresponse follow-up subsample as another phase of sample selection, where the true response distribution exists but is unknown (Kott, 1994). The literature refers to this model as the quasi-randomization model since the second phase of sample selection depends on an unknown response distribution. In the ASM application, we are therefore implementing a three-phase sample.

Since there is no workable form of variance estimator for a systematic sample, we approximate $v\left(\hat{Y}_{HT,SUB}\right)$ by treating the systematic sample as a simple random sample without replacement (Lohr, 2010). Because the variance from systematic sampling will be less than the variance from simple random sampling, our variance estimate would overestimate the true variance under 100% response in the subsample if each respondent provided complete information. We treat the random selecting of respondents from the subsample as a Bernoulli sample. The variance of $\hat{Y}_{HT,SUB}$ is given by

$$v(\hat{Y}_{HT,SUB}) = \sum_{i\in F_t} \frac{\left[1 - \pi_{i,A}(1 - \pi_{i,R})\,\pi_{i,K}\,\pi_{i,F}\right]}{\pi_{i,A}{}^2\,\pi_{i,K}{}^2\,\pi_{i,F}{}^2}\, y_i{}^2 + \sum_{i\in F_t,\ j\in F_t,}\ \sum_{i\neq j} cov(C_i I_{i,K}, C_j I_{j,K})$$

We approximate[2] the variance estimate of $\hat{Y}_{HT,SUB}$ as

$$v(\hat{Y}_{HT,SUB}) \approx \sum_{i\in F_t} \frac{\left[1 - \pi_{i,A}(1 - \pi_{i,R})\,\pi_{i,K}\pi_{i,F}\right]}{\pi_{i,A}{}^2\,\pi_{i,K}{}^2\,\pi_{i,F}{}^2}\, y_i{}^2 - \frac{(K-1)}{(nr_{NR}-1)} \sum_{i\in F_t,\ J\in F_t,}\ \sum_{i\neq j} \frac{y_i\, y_j}{\pi_{i,A}\,\pi_{jA}\,\pi_{i,K}\,\pi_{j,K}\,\pi_{i,F}\,\pi_{j,F}}$$

When $K = 1$ (100% follow-up), the covariance term reduces to 0. However, we must also account for the covariance between $\hat{Y}_{HT,ORIG}$ and $\hat{Y}_{HT,SUB}$. An unbiased estimator of this covariance is given by

$$cov(\hat{Y}_{HT,ORIG}, \hat{Y}_{HT,SUB}) = \frac{-\pi_{i,R}\,(1 - \pi_{i,R})}{2}\left(\sum_{i\in R_{ORIG}} \frac{y_i{}^2}{\pi_{i,A}\,\pi_{i,R}} + \sum_{i\in F_t} \frac{y_i{}^2}{\pi_{i,A}\,(1 - \pi_{i,R})\,\pi_{i,K}\,\pi_{i,F}}\right)$$

See Cochran (1977), Särndal et al. (1992), and Lahiri (2012).

This variance estimator assumes a three-phase sample with a Horvitz-Thompson estimator that uses the quasi-randomization estimator in the subsample to account for remaining nonresponse. It is appropriate for the ASM design and the proposed estimator. In a Census setting, the first component would be greatly reduced (or nonexistent), and the $v(\hat{Y}_{HT,SUB})$ will change with a different nonresponse adjustment procedure (e.g., ratio estimation, calibration estimation). For the case of 100% follow-up and full response, our estimator of the variance reduces to the standard Poisson sampling variance estimator of the Horvitz-Thompson estimator.

3. Case Study

3.1. Background on the Annual Survey of Manufactures (ASM)

The purpose of the ASM is to produce "sample estimates of statistics for all manufacturing establishments with one or more paid employee" (http://www.census.gov/manufacturing/asm/). The ASM collects general manufacturing statistics including the number of employees, total payroll, and shipments. The ASM is a Pareto-PPS sample of approximately 50,000 establishments selected from a universe of 328,500 manufactures. Approximately 20,000 establishments are included with certainty (probability =1), and the remaining establishments are selected with probability proportional to a composite measure of size (≈30,000 establishments). Selected units are in the sample for the four years between censuses. Although the ASM uses a Pareto sample, the publication variance estimates use the Poisson sampling variance formula. Consequently, we use Poisson sampling formula with the initial sample in our simulation study.

The ASM uses a difference estimator to reduce the sampling variance caused by probability sampling. There is a strong correlation between current-year-values and the previous Census values for manufacturing establishments, which makes a difference estimator ideal for reducing sampling variance when compared to the simple Horvitz-Thompson estimator (Särndal et al., 1992). Difference estimates are computed at the establishment level

---

[2] As part of the simple random sample without replacement approximation of the systematic sample, the joint conditional probability of being subsampled given initial nonresponse to the sample $(\pi_{ij,K})$ was approximated as the product of the individual conditional probabilities $(\pi_{i,K}\ \pi_{j,K})$, for computational ease, where convenient.

(http://www.census.gov/manufacturing/asm/how_the_data_are_collected/index.html). To reduce respondent burden, sampled units whose measure of size falls below a predetermined threshold *are not* mailed a questionnaire. Instead, their complete records are imputed using a combination of administrative data substitution and model imputation. Likewise, the ASM imputes a complete record for unit nonrespondents. For more details on the ASM design and imputation procedures, see the ASM website (http://www.census.gov/manufacturing/asm/).

Because the ASM questionnaire is a subset of the manufactures sector questionnaire in the Economic Census, the ASM is often used to pretest new processing or data collection procedures[3]. In fact, it is an excellent vehicle for a practical field test. With the ASM and the Economic Census, implementing a probability subsample of nonrespondents for follow-up represents a major procedural change. The ASM phone follow-up procedures focus on obtaining respondent data from the largest or most difficult to impute cases. The certainty units are thus a higher priority for phone follow-up. Because a given company can comprise several establishments, multi-unit (MU) establishments can be designated for phone follow-up, as company data may need to be allocated to the establishment level. All the remaining nonresponding cases receive some form of reminder, but the noncertainty single unit (SU) establishments are the least likely to receive a phone follow-up.

While a field test of the new sampling and data collection procedures is highly desirable before implementing a large-scale change in the 2017 Economic Census, the ASM has reliability requirements. Thus, we are constrained to subsampling only the noncertainty single units (SU) establishments. These cases represent approximately 5% of the total expected value of shipments and 20% of the ASM sample but also contribute approximately 11% to the total variance.

Restricting the universe for subsampling to this small subpopulation helps ensure that the required ASM reliability requirements for publication are not adversely impacted by the additional phase of sampling in a field test. At the same time, we greatly reduce our probability of demonstrating quality benefits of the respondent subsampling on the published measures.

In the simulation study described in the following sections, we consider a single data item (total value of shipments). We do not use the difference estimator implemented in the ASM. The adjustment weighting model in the simulation is equivalent to a mean imputation model, which is less precise than the ratio imputation models used in the ASM production. Not only are we constrained to a very restricted setting for subsampling, but we are also using a different estimator. For these reasons, the results presented below are not directly applicable to the ASM.

3.2 Simulation Study Parameters

The ASM conducts a mail follow-up of nonrespondents at fixed calendar dates. There are four rounds of mail-out follow-up. For the first, third, and fourth rounds, the nonrespondents are sent reminder letters, and the language in the letters strengthens at each round. The second round is a re-mail of the appropriate blank form along with the reminder letter. Bulk delivery mail-out costs are included in the survey's annual budget. Mail-out costs *per unit* are not readily available – and are not really meaningful (bulk mail is fixed price, so the cost per unit can actually increase when few letters or packages are sent.). Phone costs per unit are equally difficult to obtain, as the collected costs do not differentiate between outgoing and incoming calls, and it is impossible to distinguish between nonresponse follow-up calls and respondent question calls. Lastly, for large MU establishments, the ASM follow-up procedure is performed jointly with Census Bureau's Company Organization Survey (COS) follow-up, adding another layer of difficulty to obtain cost per unit estimates. In addition, our study only focuses on single noncertainty units, which are not budgeted separately from the rest of the ASM sample.

We obtained costs per unit for the simulation by proportionally reducing the survey's aggregate cost estimates to create totals for single noncertainty units. Mail out cost ($2.75/form, $0.75/letter), mail response cost ($0.90), and phone response cost ($5.60) are all "on average" estimates that were obtained from ASM survey experts. We estimated the survey cost for mail out and follow-up (total budget) for single noncertainty units as $68,818.

---

[3] In years ending in 2 and 7, the ASM is absorbed into the Economic Census; in other years, it is a fixed sample chosen without replacement from the recent Economic Census.

Similarly, the ASM does not track response rates for single noncertainty units. We estimated the number of responding sample units for a contact attempt to be the difference between the number of mailed packages for two consecutive contact attempts. We assume a missing-at-random (MAR) response mechanism at each follow-up round. Table 1 provides the response propensities used in our simulation, estimated as the proportion of responding sample units to the total number of mailed packages for a contact attempt. Phone and mail response propensities were estimated by splitting a contact attempt response propensity by the overall proportion of records that responded by mail and phone, respectively.

Table 1: Estimated Response Propensities for Single Noncertainty Units in the ASM

	Mail Response Propensity	Phone Response Propensity	Overall Response Propensity	Nonresponse Propensity
Initial Mail	0.21	0.14	0.35	0.65
1st Follow-up	0.18	0.12	0.30	0.70
2nd Follow-up	0.12	0.08	0.20	0.80
3rd Follow-up	0.11	0.07	0.18	0.82
4th-10th Follow-up	0.02	0.01	0.03	0.97

Table 2 provides summary statistics on the population single noncertainty units in the ASM sample. Value of shipments has a mildly skewed distribution. Although the units are not homogeneous in size (the sampling weights do vary), the range of weights is not large compared to other Economic programs. Sorting the nonresponding units by weight will induce a monotone trend – an advantage for systematic sampling – but the trend is not very steep. Any selected subsample from this population will be somewhat homogeneous, especially under a MAR response mechanism.

Table 2: ASM Population Characteristics (Single Noncertainty Units)

Variable	Mean	Q_{25}	Median	Q_{75}	Range
Value of Shipments	10,421	2,407	6,036	12,599	338,530
Sampling Weight	4.63	1.68	2.85	5.49	19
Weighted Value of Shipments	24,896	10,497	17,801	29,385	752,533

3.3. Simulation Procedure

We applied the simulation procedure described in Section 2 to the fully-imputed 2010 ASM sample and artificially induced initial unit nonresponse using empirical probabilities from Table 2 in 5,000 independent replicates. For our study, we ignore the unit and item nonresponse present in the ASM population, treating the original imputed values as reported.

In each replicate, we select four independent systematic subsamples of nonrespondents with subsampling intervals (*K*) of 1, 1.5, 2 and 4, which represent 100% follow-up, two-thirds subsample, half-subsample and quarter-subsample, respectively. After selecting a subsample, at each round of follow-up, we randomly assign the response status to the units that have still not responded. In a given round t = (1-10), units either respond by phone with probability $\pi_{p,t}$, respond by mail with probability $\pi_{m,t}$, or do not respond with probability $1 - \pi_{p,t} - \pi_{m,t}$. After round three, $\pi_{m,t}$ is always equal to .02 and $\pi_{p,t}$ is always equal to .01 (historically, there is almost no additional response after the third follow-up in the ASM).

Within replicate, we compute the running cost and estimate the population total and variance of our estimator after each stage of follow-up (K =1, 1.5, 2, 4). With 100% follow-up (K=1), there are four complete rounds of follow-up, with a slight budget balance left for a fifth incomplete round (not all planned units for follow-up were contacted). With the subsamples, we stop the follow-up after ten rounds, with some remaining funds[4].

We conducted the simulation under two different scenarios (Scenario 1 and Scenario 2). In Scenario 1, the systematic subsample is selected *before* follow-up begins, so that all rounds of follow-up are applied to the probability subsample. In Scenario 2, all nonresponding units are mailed a reminder letter, then a systematic subsample is selected. As a result, at least one follow-up contact effort is made for all nonresponding units, not just the subsampled units.

In both scenarios, we track running cost (1) and compute the average variance at each follow-up round (2) over the 5,000 replicates. In addition, we compute the relative bias of $\widehat{Y}_{HT,K,t}$ (the Horvitz-Thompson estimate with K=*k* at round *t* of follow-up) as

$$\text{Relative Bias: } \frac{1}{5{,}000}\sum_{s=1}^{5{,}000}\frac{\widehat{Y}_{HT,K,s,t}-Y}{Y}$$

where Y= the “true” total value of shipments in the single noncertainty unit ASM population.

The Appendix presents selected results from the simulation study under both scenarios. In both scenarios and for all subsampling rates, the Horvitz-Thompson estimates are essentially unbiased, as expected with the induced MAR response mechanism.

4.Simulation Results

4.1 Subsampling Nonrespondents Before Follow-Up (Scenario 1)

In the first scenario, only subsampled nonrespondents receive any follow-up. This design maximizes cost savings. Figure 1 graphs the running cost against follow-up attempts. The red line represents the current estimated budget for follow-up of single noncertainty units. Follow-up attempt zero is the initial cost for mail out and response. Not surprisingly, the larger the sampling rate (highest being 100% follow-up, K = 1), the more expensive the follow-up.

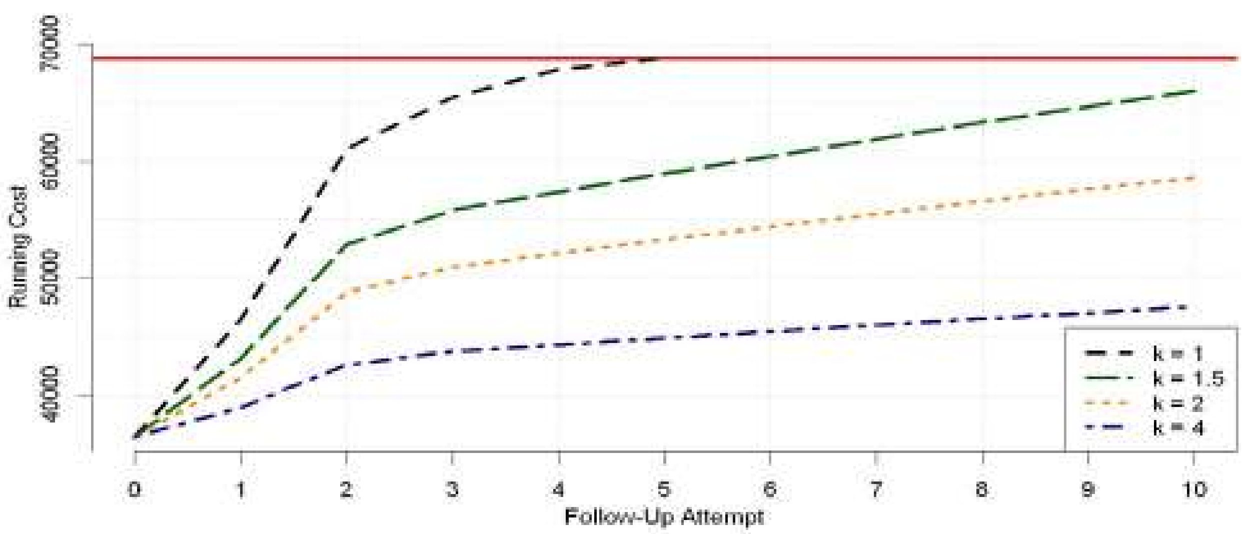

Figure 1: Follow-Up Attempts vs. Average Running Cost (Scenario 1)
Source: 2010 ASM Tabulation Microdata (SU Noncertainty Cases) with Artificially Induced Nonresponse

---

[4] Under Scenario 2, when K=1.5, the total budget was exhausted during but before completion of the tenth round, for nearly all of the 5,000 replicates. For some replicates, ten complete rounds of follow-up were conducted. However, for one replicate, funds were exhausted during the ninth round.

Figure 2 graphs the unweighted unit response rates for each subsampling rate, where the unweighted response rate is the number of responding units divided by the original number of single noncertainty units in the initial sample. The red line represents the overall unweighted response rate for ASM, roughly 72 percent. As expected, only the 100% follow-up plan comes close to achieving this rate. Since the initial response rate is roughly 35 percent, subsampling the remaining 65 percent of sampling units limits the maximum achievable unit response rate. For example, *K*=4 takes a quarter sample of 65 percent of the initial nonrespondents. The largest possible unit response rate for *K*=4 is therefore approximately 51 percent.

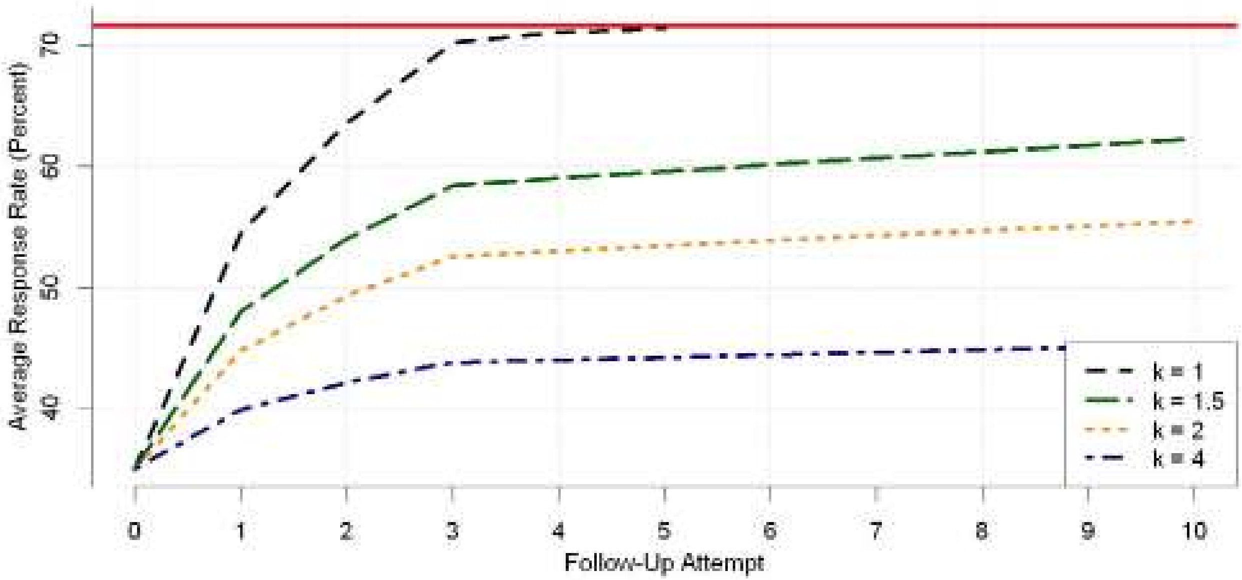

Figure 2: Follow-Up Attempts vs. Average Unweighted Response Rate (Scenario 1)
Source: 2010 ASM Tabulation Microdata (SU Noncertainty Cases) with Artificially Induced Nonresponse

Recall though that the single noncertainty units represent a portion of the complete survey sample units (about twenty percent). Consequently, a decreased unit response rate caused by subsampling nonrespondents would have an impact on the ASM unit response rates.

The Appendix presents the relative bias computed for totals shipments after each follow-up by sampling rate. All estimates are essentially unbiased, with relative biases of less than 0.07 percent. Of course, the simulation design employs a MAR response mechanism, so this unbiasedness is more likely a reflection on the simulation's correct implementation than a general commentary on the estimator used.

Figures 3 and 4 plot the average variance estimates against follow-up round for each sampling interval. The red line marks the estimated Poisson sample variance for single noncertainty units under full response. Recall that the ASM imputes nonrespondents with administrative data substitution and ratio imputation. We did not use similar (more precise) imputation procedures. The ASM difference estimates are also more precise than our Horvitz-Thompson estimates. Thus, the ASM production variance estimates should be, and are, much smaller than our l00% follow-up variance estimates.

Alone, systematic subsampling increases each variance estimate for the nonrespondents subsample by K^2. With *K*=4, this an unacceptably large increase. Moreover, the remaining nonresponse in the subsamples increases the variance and is reflected by the large and quite variable adjustment cell weighting factors. This phenomenon is further discussed in Section 4.2.

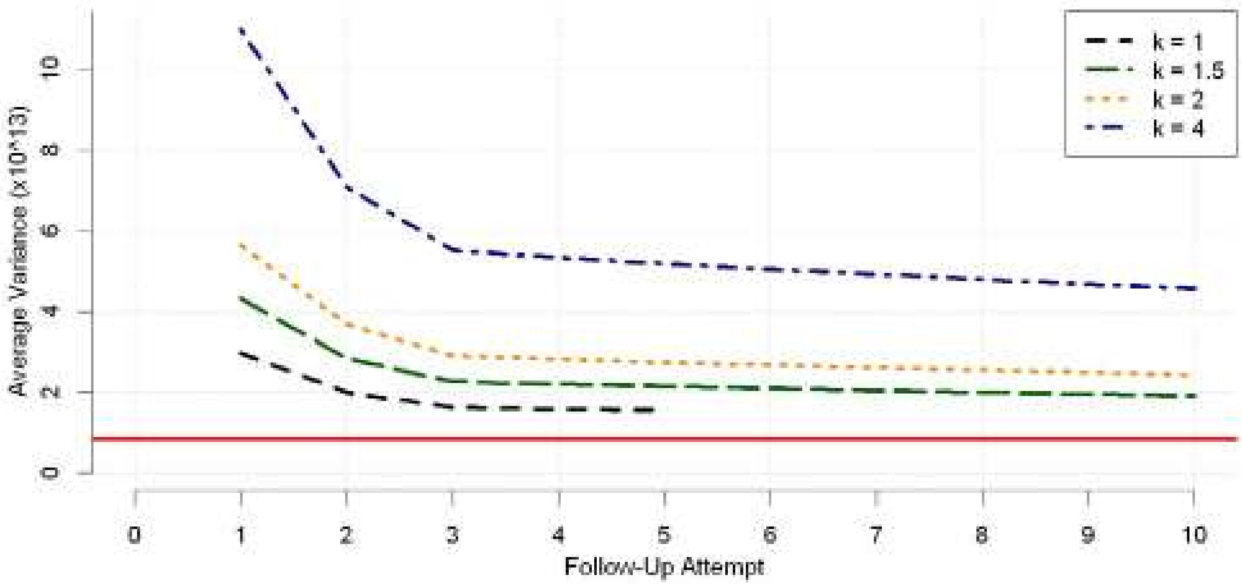

Figure 3: Follow-Up Attempts vs. Average Estimated Variance (Scenario 1)
Source: 2010 ASM Tabulation Microdata (SU Noncertainty Cases) with Artificially Induced Nonresponse

Overall, these are not unexpected results. Subsampling nonrespondents increases the sampling variance. Conversely, selecting a subsample of nonrespondents for follow-up greatly reduces the cost and the smaller number of eligible units allows for increased number of rounds of follow-up for the same budget. Ultimately, the decision to subsample depends on how much precision in terms of sampling variance can be sacrificed for lower costs.

Figure 4 provides a tool for this decision, graphing the relationship between cost, variance, and follow-up round. Each circle represents an estimate of cost and variance after a follow-up round. For example, when K=2, the top left yellow circle represents an average total cost of $41,417 with an average estimated variance of roughly 5.64 X 10^{13} after one round of follow-up. By the third follow-up, with K=2, the variance estimate has nearly halved for the additional $10,000, but is still considerably larger than the variance after three rounds of follow-up for either K=1 or K=1.5.

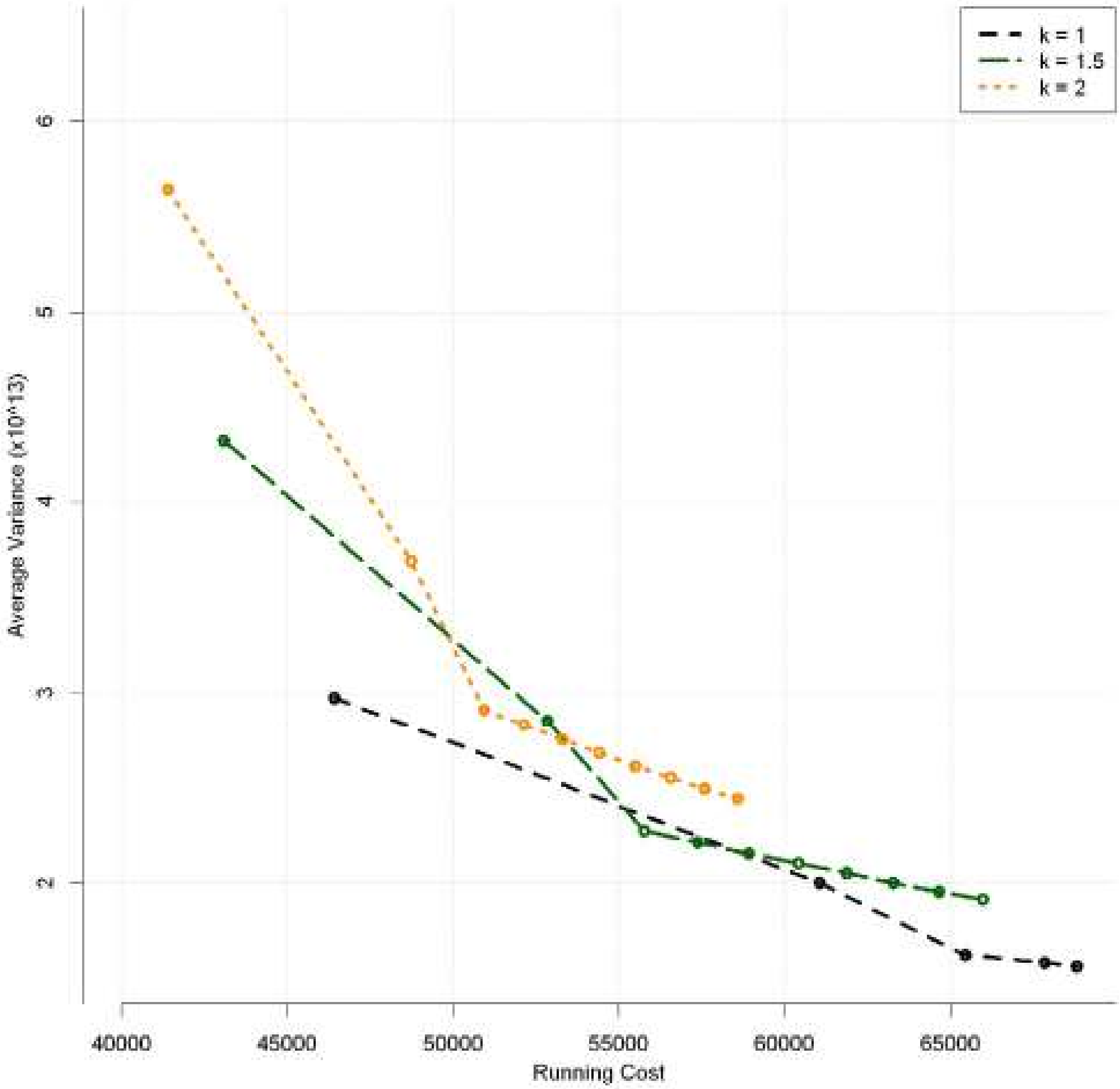

Figure 4: Running Cost vs. Average Estimated Variance by Follow-Up Round (Scenario 1)
Source: 2010 ASM Tabulation Microdata (SU Noncertainty Cases) with Artificially Induced Nonresponse

Figure 4 show how many additional rounds of follow-up are needed to achieve the comparable levels of variance under the alternative nonrespondent subsampling designs. With our propensity model, the later rounds of follow-up yield minimal improvements in variance because the response propensity is very close to zero after the third follow-up. We use this graph to compare the effects of alternative sampling intervals on variance. However, one can also use the graph to determine the additional variance that would be induced in the estimate if planned follow-up rounds were "cut short" due to budget constraints.

4.2 Subsampling Nonrespondents after the First Round of Follow-up (Scenario 2)

In an adaptive design framework, a data collection procedure should not change until it reaches a "phase capacity." Clearly, the ASM follow-up has not reached this state before the first follow-up, which is very beneficial in increasing response. To keep the benefit of a complete recontact of nonrespondents at the first follow-up, we modified the subsample design to begin subsampling nonrespondents after completing the first round of contact. Subsampling before the second round fits well into our cost model because the second round is a re-mail of the initial survey package, which is considerably more expensive than mailing a reminder letter. The costs for K=1 only differ minutely from Scenario 1 due to random fluctuations, but costs are somewhat increased for the other subsampling intervals as one round of complete follow-up is now conducted before subsampling. Thus, the cost figures are virtually identical after round 1 for all values of K, but diverge after round 2.

Figure 5 graphs the relationship between cost, variance, and follow-up round under this scenario, using the same axis as Figure 4 for comparison. Since subsampling in this scenario does not begin until the first round of follow-up is completed, the first circle from left to right on Figure 5 represents the running cost and average variance after the *second* round of follow-up.

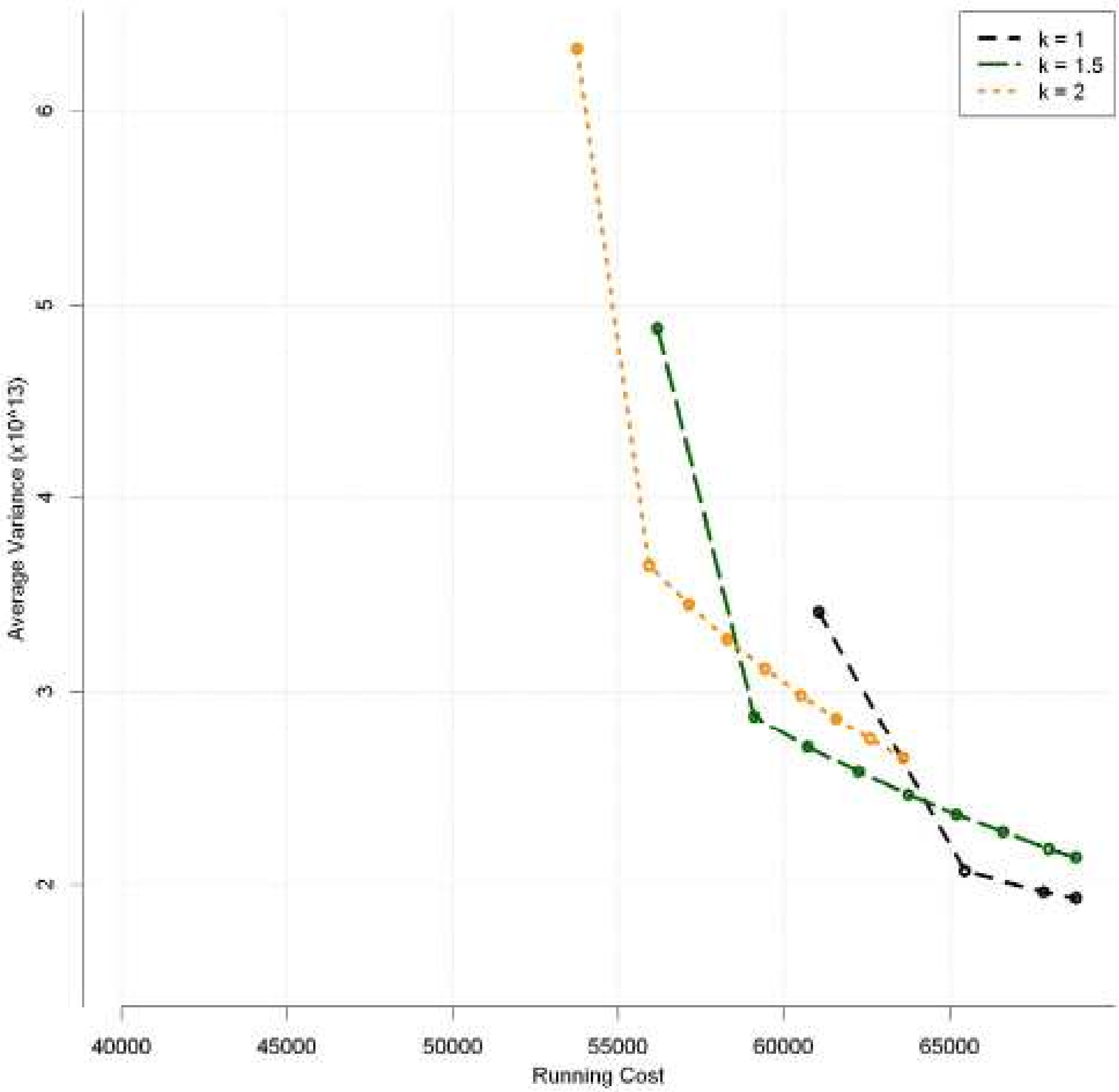

Figure 5: Running Cost vs. Average Estimated Variance by Follow-Up Round (Scenario 2)
Source: 2010 ASM Tabulation Microdata (SU Noncertainty Cases) with Artificially Induced Nonresponse

By delaying subsampling, the unit response rates increase for K=1.5, 2, and 4. Despite this increase, the estimated variance of the total estimate is notably larger under Scenario 2 as compared to Scenario 1. This result was counterintuitive. In this scenario, the "initial" sample contains more respondent units with unadjusted weights than in the first scenario. Unfortunately, the probability of responding after the first round of follow-up is not high, which increases the value of the nonresponse weight adjustment factors for the responding subsampled cases. The Appendix presents the nonresponse adjustment factors for each round of subsampling in Scenarios 1 and 2. The additional unadjusted weights did not offset the increase of the nonresponse adjustment factor applied to the smaller subsample. In addition, the ASM noncertainty units are fairly homogeneous, so the reduced subsample is more variable than a larger subsample.

Figures 6 and 7 plot the running cost and cumulative variance, respectively, for Scenarios 1 and 2 again the total number of follow-ups for K=1 and K=2.

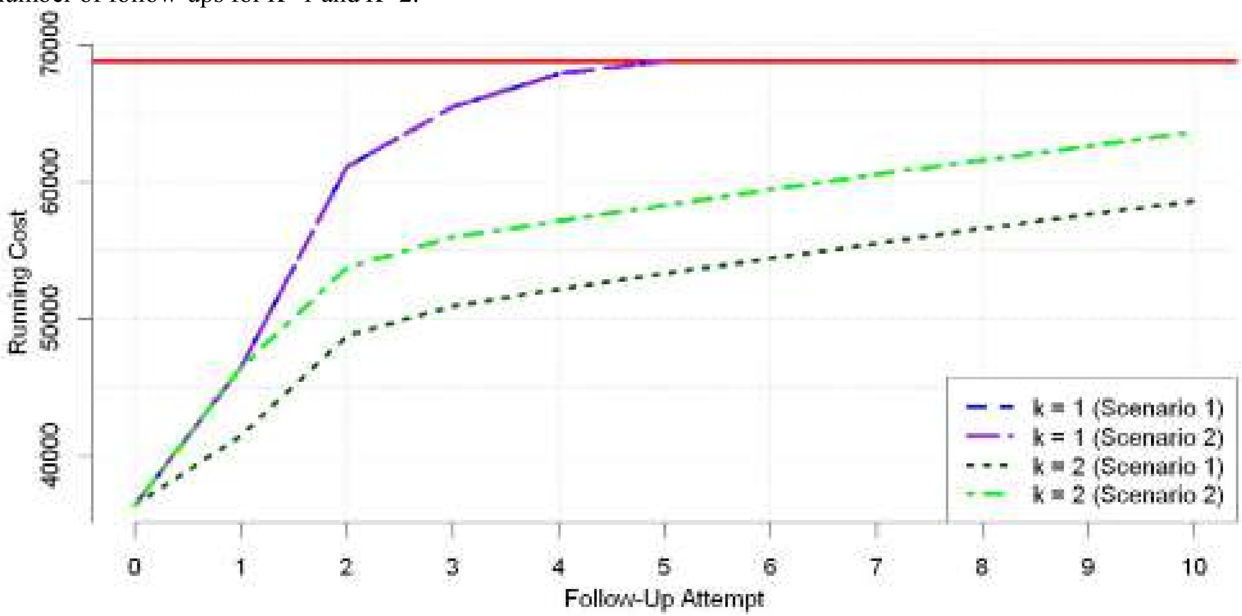

Figure 6: Side-by-Side Comparison of Running Cost (Scenarios 1 and 2).
Source: 2010 ASM Tabulation Microdata (SU Noncertainty Cases) with Artificially Induced Nonresponse

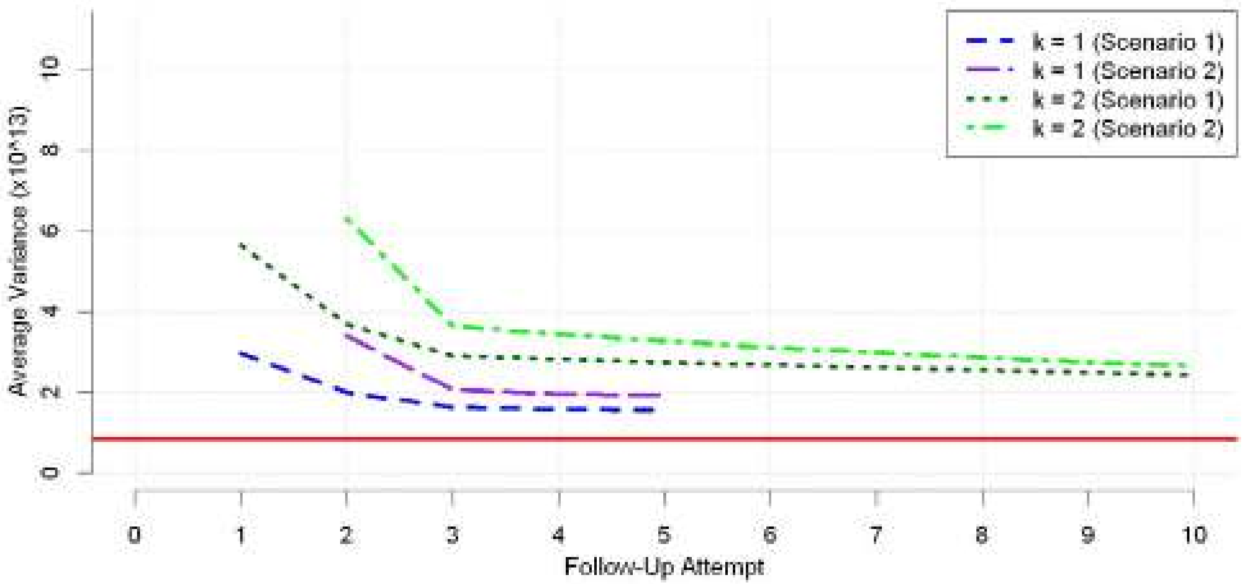

Figure 7: Side-by –Side Comparison of Average Variance for Scenarios 1 and 2
Source: 2010 ASM Tabulation Microdata (SU Noncertainty Cases) with Artificially Induced Nonresponse

The running costs are the same in both scenarios for *K*=1 (100% follow-up), but differ for *K*=2 after the first follow-up round. Both *K*=2 subsamples complete ten rounds of follow-up for less cost than four rounds of follow-up with *K*=1. After seven rounds of follow-up in both subsamples, the difference in average variance is fairly trivial. In using the more adaptive design, the slight increase in sampling variance costs an additional $5,000 (about seven-percent of the allocated budget), with potential quality gains that could be realized by using different collection procedures on the smaller subsample of nonrespondents.

5. Conclusion

In this paper, we present an approach for evaluating alternative allocation rates for sampling nonrespondents, subsampling considering both cost and variance. Using a simulation approach, we examined the relationship between cost and variance for various allocation rates. Unfortunately, our approach does not account for other quality benefits of subsampling, such as reduced estimation bias from the more representative sample.

We examine the cost and variance trade-offs in systematically subsampling nonrespondents with one estimator and a single item, where the subsampled population is fairly homogeneous. The estimator is sample-based and does not incorporate any auxiliary information. Moreover, our variance estimator is conservative and could be a large upper bound on the true variance from a systematic subsample. Even so, it would be difficult to advocate selecting a probability subsample of nonrespondents from this survey subpopulation with this estimator if the decision were based on variance considerations alone. However, cost is always a factor, and "fixed" budgets are becoming quite elusive. Our simulation approach gives a tool for assessing the cost and variance trade-offs for a given design at any round of follow-up (see Figures 4 and 5).

It is important to remember that there are important differences in non-response follow-up in our controlled study versus in practice. In this study, we did not assume any differentiation in quality of responses between the scenarios. However, it is likely that following up a smaller number of cases through systematic subsampling would lead to improved response quality as analyst resources target a smaller number of cases. In a true adaptive design setting, data collection procedures would change in the later rounds of follow-up, once it is determined that the current procedures are no longer working. Another factor to consider is our assignment of response propensities throughout the follow-up process. Whereas our study assumes that all response propensities are uniformly distributed across units, response is correlated with unit size in business surveys. In this setting, the subsampling procedure might benefit from an initial stratification of nonrespondents by unit size, then a more proportional allocation focusing on smaller units.

We present an analysis approach. Before making specific recommendations for the survey, we need to investigate other estimators that use auxiliary data in the nonresponse adjustment. Bechtel and Thompson (2013) show very promising results with ratio estimation instead of adjustment cell weighting, and we should explore that in future simulations. Another possibility is to use auxiliary information to predict the probability of response to any potential nonresponse follow-up and then re-weight the respondents based on their predicted probability of response. Using random auxiliary variables in this way would not increase the variance of the total estimator, provided that there is no misspecification in the nonresponse model (Beaumont, 2005). In fact, Little and Vartivarian (2005) argue that including such auxiliary variables related to the variable of interest should reduce any nonresponse variance (as cited in Beaumont, 2005). The first investigation is a simple change to our simulation programs; the second is more involved.

Of course, the potential benefits of either 100% follow-up or selecting a probability subsample of nonrespondents depend on the structure of the population under study. Conducting complete follow-up on populations is expensive and may lead to biased samples if certain types of units are more likely to respond than other types. That said, systematically subsampling populations that do not exhibit a clear monotone trend simply increases the sampling variance (Lohr, 2010). Using an adaptive design to take the route that achieves the most benefit at any time allows us to come closest to achieving an optimal design.

Acknowledgments

The authors wish to thank Eric Fink, Xijian Liu, Broderick Oliver, Robert Struble, and Edward Watkins III for their review and comments. We also wish to thank Michelle Vile Karlsson and Michael Zabelsky for providing ASM cost data for our study.

References

Beaumont, J-F. (2005). On the Use of Data Collection Process Information for the Treatment of Unit Nonresponse Through Weight Adjustment. *Survey Methodology*. **31**, 227-231.

Bechtel, L., & Thompson , K. J. (2013). Optimising unit nonresponse adjustment procedures after subsampling nonrespondents in the Economic Census.

Cochran, W. G. (1977). Sampling Techniques. (3rd ed.). New York, NY: John Wiley & Sons.

Groves, R. M. and Heeringa, S. G. (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*(3), 439-457.

Kalton, G. and Flores-Cervantes, I. (2003). Weighting Methods. *JOS*, **19**, pp. 81-97.

Kott, P. (1994). A Note on Handling Nonresponse in Sample Surveys. *Journal of the American Statistical Association*, **89**, pp. 693 –696.

Lahiri, P. (2012). Chapter 2: Framework for Probability Sampling [Course Notes.]. Personal Collection of Lahiri, P, JPSM, University of Maryland, College Park, MD.

Little, R.J., and Vartivarian S. (2005). Does weighting for nonresponse increase the variance of the sample means? *Survey Methodology*, **31**, 161-168.

Lohr, S. L. (2010). Sampling: Design and Analysis. (2nd ed.). Boston, MA: Brooks/Cole.

Särndal, C., Swensson, B., and Wretman, J. (1992). Model Assisted Survey Sampling. (1st ed.). New York: Springer-Verlag.

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

Appendix: Simulation Study Results

	Follow-up Round	Running Cost		Average Variance $\times 10^{13}$		Response Rate		Relative Bias		Nonresponse Adjustment Factor (*Multiply by K)*	
		S1[5]	S2[6]	S1	S2	S1	S2	S1	S2	S1	S2
Subsampling Factor K = 1	0	36,406	36,406			35.0%					
	1	46,427	46,429	2.97		54.5%	54.5%	-0.02%		3.33	
	2	61,068	61,066	2.00	3.41	63.6%	63.6%	0.00%	-0.01%	2.27	5.00
	3	65,464	65,462	1.62	2.07	70.2%	70.2%	0.00%	0.01%	1.85	2.91
	4	67,858	67,856	1.58	1.96	71.1%	71.0%	0.00%	-0.01%	1.80	2.75
	5	68,819	68,819	1.56	1.93	71.4%	71.4%	0.00%	0.00%	1.76	2.61
Subsampling Factor K = 1.5	0	36,406	36,406			35.0%					
	1	43,085	46,425	4.32		48.0%	54.5%	0.01%		3.33	
	2	52,846	56,185	2.85	4.88	54.1%	60.6%	-0.01%	0.02%	2.27	5.00
	3	55,777	59,116	2.27	2.87	58.4%	64.9%	0.00%	0.03%	1.85	2.91
	4	57,374	60,713	2.21	2.71	59.0%	65.5%	0.00%	0.02%	1.80	2.75
	5	58,922	62,262	2.15	2.58	59.6%	66.1%	0.00%	0.02%	1.76	2.61
	6	60,424	63,763	2.10	2.46	60.2%	66.7%	0.00%	0.03%	1.72	2.49
	7	61,882	65,220	2.05	2.36	60.7%	67.2%	0.00%	0.03%	1.68	2.39
	8	63,294	66,634	2.00	2.27	61.2%	67.7%	-0.01%	0.03%	1.65	2.29
	9	64,665	68,004	1.95	2.18	61.8%	68.3%	-0.01%	0.03%	1.62	2.20
	10	65,995	68,817	1.91	2.14	62.3%	68.6%	0.00%	0.03%	1.59	2.13
Subsampling Factor K = 2	0	36,406	36,406			35.0%					
	1	41,417	46,427	5.64		44.8%	54.5%	0.00%		3.33	
	2	48,737	53,745	3.69	6.32	49.3%	59.1%	0.02%	-0.04%	2.27	5.00
	3	50,935	55,942	2.91	3.65	52.6%	62.3%	0.01%	-0.01%	1.85	2.91
	4	52,132	57,139	2.83	3.45	53.0%	62.8%	0.01%	-0.01%	1.80	2.75
	5	53,293	58,301	2.75	3.27	53.5%	63.2%	0.01%	-0.01%	1.76	2.61
	6	54,419	59,428	2.68	3.12	53.9%	63.6%	0.01%	-0.01%	1.72	2.49
	7	55,511	60,520	2.61	2.98	54.3%	64.0%	0.01%	-0.02%	1.68	2.39
	8	56,571	61,581	2.55	2.86	54.7%	64.4%	0.00%	-0.02%	1.65	2.29
	9	57,598	62,609	2.49	2.75	55.1%	64.8%	0.01%	-0.02%	1.62	2.20
	10	58,595	63,606	2.44	2.65	55.4%	65.2%	0.01%	-0.02%	1.59	2.13
Subsampling Factor K = 4	0	36,406	36,406			35.0%					
	1	38,909	46,432	10.98		39.9%	54.5%	-0.02%		3.33	
	2	42,569	50,089	7.10	12.33	42.2%	56.8%	0.00%	0.04%	2.27	5.00
	3	43,669	51,188	5.52	6.87	43.8%	58.4%	-0.04%	0.04%	1.85	2.91
	4	44,268	51,787	5.35	6.46	44.0%	58.7%	-0.05%	0.03%	1.80	2.75
	5	44,848	52,367	5.19	6.10	44.2%	58.9%	-0.06%	0.03%	1.76	2.61
	6	45,411	52,930	5.05	5.79	44.4%	59.1%	-0.06%	0.03%	1.72	2.49
	7	45,957	53,476	4.92	5.52	44.6%	59.3%	-0.06%	0.03%	1.68	2.39
	8	46,488	54,006	4.80	5.27	44.8%	59.5%	-0.06%	0.04%	1.65	2.29
	9	47,001	54,520	4.68	5.05	45.0%	59.7%	-0.06%	0.03%	1.62	2.20
	10	47,500	55,018	4.57	4.86	45.2%	59.9%	-0.06%	0.04%	1.59	2.13

[5] S1 refers to Scenario 1. For additional information regarding Scenario 1, refer to p. 10.

[6] S2 refers to Scenario 2. For additional information regarding Scenario 2, refer to p. 10.