Responsiveness and Representativeness in an Establishment Survey of Manufactures

Eric B. Fink[1] and Joanna Fane Lineback
U.S. Census Bureau, Washington, D.C. 20233

Abstract
Response rates, because of their ease of calculation and understanding, traditionally have been used as data-collection-quality metrics. However, research has cautioned against solely relying on response rates, as survey programs' aims to increase these rates may lead to increasing the likelihood of biasing survey estimates (Groves, 2006). R-indicators have been proposed as a corresponding measure that can give insight into the data collection process that response rates alone cannot explain (Schouten and Cobben 2007). In this paper, we calculate traditional response rates and R-indicators for the 2011 Annual Survey of Manufactures and demonstrate that when used in conjunction with each other they can give a more complete picture of the data collection process, particularly the nonresponse follow-up. In particular, we show that despite increasing response rates during the nonresponse follow-up, representativeness across important design variables such as establishment size decreases, owed in part, we hypothesize, to concentrating follow-up on those establishments expected to contribute the most to total estimates. This lack of representativeness is a possible source of bias in resulting survey estimates if nonresponse adjustments do not correct for over or underrepresented areas. We discuss the tradeoff of reducing sampling variability versus reducing nonresponse bias. Further, we incorporate associated costs into our analysis, and discuss how these cost/quality indicators can be used in conjunction with data quality metrics to provide a more complete picture of the efficacy of the survey process.

Keywords: Response Rates, R-indicators, Annual Survey of Manufactures, Establishment Survey

1. Introduction

As many surveys in the Economic Directorate of the U.S. Census Bureau rely on response rates as the metric for understanding data collection performance, and as research has demonstrated that solely relying on response rates may bias survey estimates (Groves, 2006), we decided to explore R-indicators as a complementary metric to compute during and after data collection to help make decisions about the nonresponse follow-up process. R-indicators may provide a useful tool to evaluate how closely the sample obtained at the end of (or during) the collection process mirrors the sample initially drawn. We will not argue that R-indicators should be used in lieu of response rates, but demonstrate that when used in conjunction with response rates can provide a more complete picture of the data collection process.

We demonstrate the utility of examining both response rates and R-indicators by developing a profile of respondents at key stages of data collection using frame and paradata variables captured in the 2011 ASM. We first look at unit and item-level response rates in the 2011 Annual Survey of Manufactures (ASM). Because the ASM imputes missing data using the respondent pool, we also look at response behavior by how businesses are organized, as single or multi-unit enterprises (see below for a description). We do this because larger companies have an impact on confidence interval width for point estimates, and the ASM is a longitudinal survey for which estimates of change are most important. If larger companies are targeted more acutely in nonresponse follow-up, decreasing the representativeness of smaller companies could introduce bias in these estimates. Accordingly, we look to R-indicators to assess how representative the final sample was relative to the sample initially drawn. The model-based estimates generated to calculate these R-indicators were developed using variables important in sample design for the ASM.

[1] Economic Statistical Methods Division (eric.fink@census.gov) This report is released to inform interested parties of research and to encourage discussion. Any views expressed on methodological or operational issues are those of the author and not necessarily those of the U.S. Census Bureau.

These data used to conduct our analysis come from the Business Register (BR). The BR is a centralized business database where information for enterprises, establishments, and other administrative data are stored. Frame data include industry classification, geographic data, and type of establishment while paradata include response flags, time of form check-in, and mail-out date. Additionally, we have cost data that includes costs associated with the initial survey mail-out operation, as well as mail and telephone follow-up.

For this paper, we define an establishment as a single physical location where business is conducted or where services or industrial operations are performed. Further, the terms “establishment” and “unit” are used interchangeably in this paper. Finally, we define an enterprise as a business organization consisting of one or more domestic establishments that were specified under common ownership or control. The enterprise and the establishment are the same for single-unit (SU) organizations. Each multi-unit (MU) company forms one enterprise.

In this paper, we will discuss the following: Section 2 gives relevant background information on the ASM; Section 3 will describe the methods we used to analyze the data; Section 4 will present results. Finally, Section 5 discusses the results and future research directions.

2. Background

The ASM is a mandatory response survey that provides statistics on employment, payroll, supplemental labor costs, cost of materials consumed, operating expenses, value of shipments, value added by manufacturing, detailed capital expenditures, fuels and electric energy used, and inventories for all manufacturing establishments with one or more paid employees. In this section, we provide information on the major components of the ASM program, including sample design, data collection, including nonresponse follow-up, and estimation. For information on the ASM including historical data and forms, go to http://www.census.gov/manufacturing/asm/index.html.

2.1 Sample Design

To select the ASM sample, the manufacturing population is partitioned into two groups: establishments eligible to be mailed a questionnaire, a mail stratum, and establishments not eligible to be mailed a questionnaire, a nonmail stratum. The eligible establishments consist of larger single-location, manufacturing companies and all manufacturing establishments of multi-location companies. The nomail establishments consist of small and medium-sized, single-establishment companies based on a measure of size derived from the most recent Economic Census. Data for these ineligible establishments are estimated using information obtained from the administrative records of the Internal Revenue Service and Social Security Administration, and are included in the published ASM estimates.

The ASM mail sample includes approximately 50,000 establishments of which about 20,000 are selected with certainty, and about 30,000 are selected with probability proportional to a composite measure of establishment size. Although the nonmail stratum contained approximately 180,000 individual establishments in 2011, it accounted for less than 7 percent of the estimate for total value of shipments at the total manufacturing level. A new sample is selected at five-year intervals beginning the second survey year subsequent to the Economic Census. This information is supplemented with data for new companies from the IRS and the Census Bureau’s Report of Organization Survey (COS).

2.2 Data Collection

Data are collected annually for the ASM except for years ending in 2 and 7 when the Economic Census is conducted. The survey is establishment-based, although for a multi-establishment business the questionnaires are mailed to the business enterprise unless another reporting arrangement has been made. Respondents can choose to report by mail or electronically using either the Census Surveyor software (for multi-unit organizations) or by the Web (for single-unit organizations). In addition, respondents may fax forms and in some cases give their responses by phone. In 2011, every enterprise in the sample received a paper form[2]. All multi-units that receive a request to complete the ASM also get a request to complete the COS in the same package. Responses are due within 30 days of receiving the form. The COS is an annual mail-out/mail-back survey of selected companies with payroll, excluding companies engaged exclusively in agricultural production. The purpose of the COS is to obtain current organization and operating information on multi-establishment firms in order to maintain the BR.

Follow-up with nonresponding businesses begins approximately two months after the initial mailout and is usually in the form of a mailed letter. After the first reminder, there are three additional reminders sent, once a month, until a case is considered a delinquent nonrespondent. For some very large establishments that are deemed important for estimation purposes, follow-up may occur via telephone. Currently, data collection continues for the ASM until the project runs out of time or money.

2.3 Estimation

Most of the ASM estimates derived for the mail stratum are computed using a difference estimator. The difference estimator takes advantage of the fact that, for manufacturing establishments, there is a strong correlation among some estimates between the current year data values and the previous Economic Census values. Because of this correlation, difference estimates are considered more reliable than comparable estimates developed from the current sample data alone. The ASM difference estimates are computed at the establishment level by adding the weighted difference (between the current data and the Economic Census data) to the Economic Census data. However, some estimates are not generated using the difference estimator because the year-to-year correlations are considerably weaker. A standard linear estimator is used for these variables. Estimates are published from the 2 – 6 digit North American Industry Classification System (NAICS) level, and for the U.S. and by state.

3. Analysis

3.1 Analysis Variables

From the BR we obtained information about participation in other Census Bureau surveys, check-in dates, and the mode the respondent supplied information to the Census Bureau for the 2011 ASM. We have also obtained data indicating costs associated with initial mailing, as well as follow-up mailings and telephone costs.

3.2 Analysis Questions

Much of the research we have conducted to this point was exploratory in nature. We spent months obtaining and merging the aforementioned data and many of the initial research questions necessitate only descriptive statistics to answer. Our overarching research question is how to gain greater insight into the quality of our data collection process for the ASM. We further refined our question into several more manageable parts about business-respondent behavior. We develop the following initial questions:

1. What is the cumulative unit response rate?
2. What is the cumulative total quantity response rate?
3. How much money are we spending on each stage of data collection relative to the achieved response rate?
4. How representative is the final sample of respondents compared to the sample that was initially selected?

3.3 Limitations of the Analysis

There are limitations with respect to the cost data presented in Section 4.1. The costs here only reflect mail form and phone call costs (direct labor, overhead, and outgoing calls). At this point, we are unable to

[2] For the 2012 Economic Census if 2011 ASM responses were electronic, paper forms were not sent.

reasonably estimate cost by survey or by survey activity such as form design, sample selection, or data processing. Additionally, it is not always possible to separate ASM and COS costs because they are conducted jointly. However, as ASM is a much more involved survey instrument in that it asks much more than does COS, a reasonable simplifying assumption for this paper is that where we are given costs for both ASM and COS, a vast majority of the resources are being utilized for ASM.

4. Results

Results are given below. Both subsections 4.1 and 4.2 present results on all mailed ASM establishments in NAICS 31, the manufacturing sector. Subsection 4.1 presents results on survey response and costs, and subsection 4.2 presents results on R-indicator calculations.

4.1 Response Metrics and Costs

The unit response rate (URR) is the number of forms returned with sufficient information to be deemed a response, either by paper or electronically, as a proportion of those mailed to eligible units, as well as units of unknown eligibility. The URR covers all mailed multi-unit and single-unit establishments. Again, the URR serves as a measure of data collection performance. The curve in Figure 1 shows the approximate URR for the 2011 ASM from initial mail-out through follow-up. It shows that the response rate is increasing more quickly after the due date, then slows over time, with an overall rate just under 80%. To examine this more closely, Figure 2 shows the number of responses at 10 day intervals from initial mail-out through follow-up. It shows a sharp increase in the number of responses occurred leading up to, and shortly after the due date. Another spike occurred, in the 70-80 day interval, after the first reminder was sent. We see a bulk of the responses coming in between the 70 and 190 day intervals.

Figure 1 also shows the cumulative percentage of the mailing budget from the initial mailing through the fourth follow-up. The first percentage listed is so large because it includes the cost of printing the forms, as well as the cost of postage for the mailing plus the cost of postage on the envelope for return, in addition to early incoming and outgoing phone calls. Based on the available information, it is difficult to assess the effectiveness of follow-up mailings relative to spending. This can best be evaluated through a carefully planned experiment.

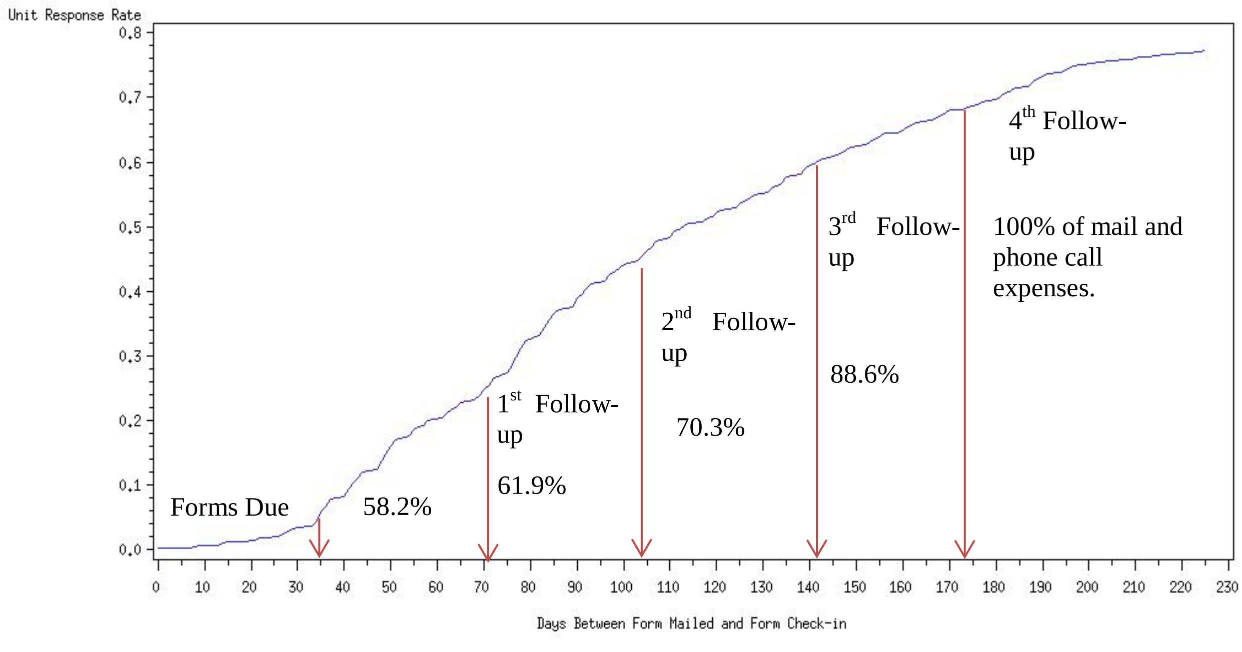


Figure 1. The unit response rate for the 2011 ASM from when forms were initially mailed to respondents. The red arrows represent mail-out dates for follow-up letters at 71, 104, 141, and 174 days for 1st, 2nd, 3rd, and 4th follow-up respectively. The percentages listed show the cumulative percent of the total mailing and telephone expenditures allocated to each stage of data collection up to, but not including, the follow-up subsequently listed.

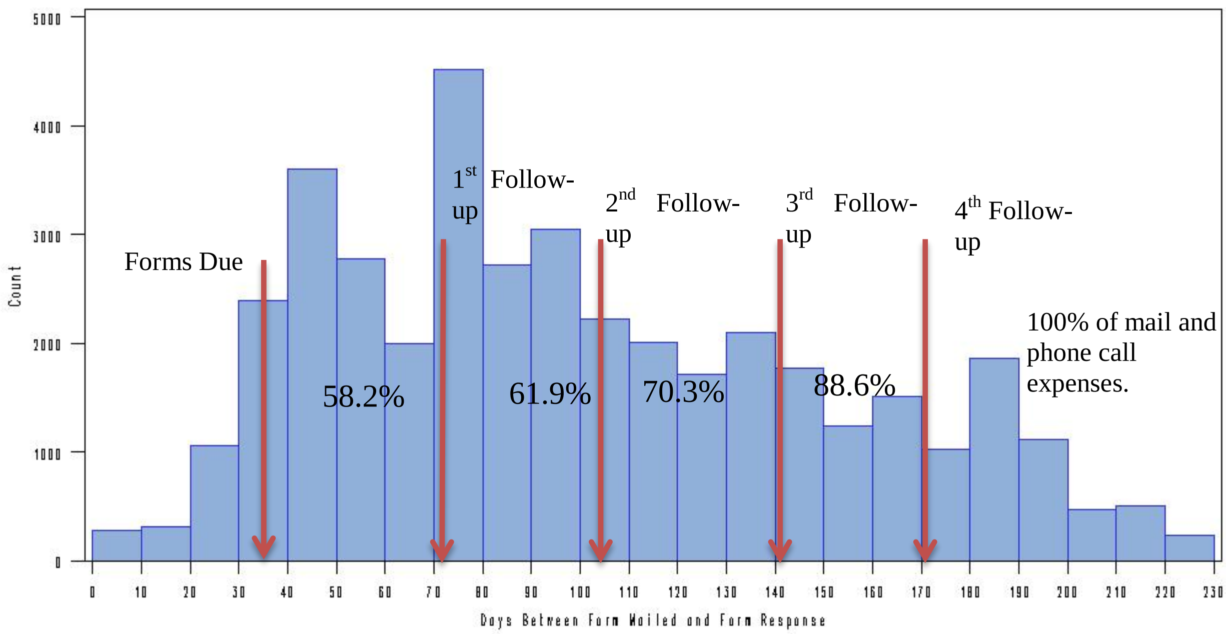


Figure 2. The number of responses at 10-day intervals for the 2011 ASM from mail-out to the end of collection. The red arrows represent mail-out dates for follow-up letters at 71, 104, 141, and 174 days for 1st, 2nd, 3rd, and 4th follow-up respectively. The percentages listed show the cumulative percent of the total mailing and telephone expenditures allocated to each stage of data collection up to, but not including, the follow-up subsequently listed.

In Figure 3, we see the URR broken out by SU and MU establishments. The red arrow indicates the third wave of nonresponse follow-up, which has more associated costs, as we saw in the prior figures. It is after this follow-up that we see a crossover, where the MU URR is higher than the SU. It is possible that during this wave there is a more effective nonresponse follow-up for the MUs. Another possibility is that about the time we see the crossover is also when companies file their Security and Exchange Commission forms. What is important in this graphic is that the slopes change, so the rate at which we are getting certain types of units is changing, an important point when planning future data collection and follow-up strategies.

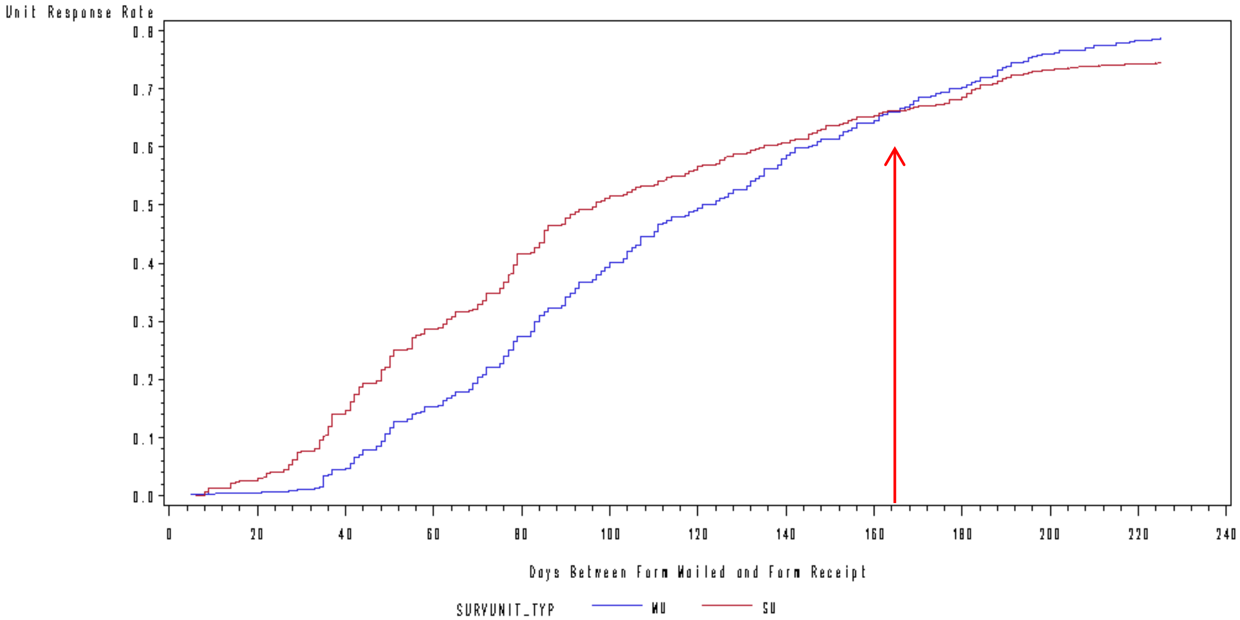


Figure 3. 2011 ASM URR by single-unit and multi-unit status.

The total quantity response rate (TQRR) is the proportion of the estimated, weighted total of data item *t* reported by the active tabulation units in the statistical period or from sources determined to be equivalent-quality-to-reported data (expressed as a percentage). See Appendix D3-B in the Census Bureau Quality Standards for more information on the TQRR (http://www.census.gov/quality/standards/Quality_Standards.pdf).

The TQRR is computed as follows:

$$TQRR = \left[\frac{\sum_{i=1}^{N_T} w_i * (r_{ti} + q_{ti}) * t_i}{\sum_{i=1}^{N_T} w_i f_i |t_i|}\right] * 100$$

Where:

- w_i is the design weight of tabulation unit *i*,
- r_{ti} is the indicator variable for reported data for tabulation unit *i* and data item *t*,
- q_{ti} is the indicator variable of "equivalent quality" data for tabulation unit *i* and data item *t*,
- t_i is the data value for unit *i*,
- f_i is the nonresponse weighting adjustment factor for tabulation unit *i*, and

N_T is the total number of eligible tabulation units.

While the URR is an unweighted measure of unit nonresponse, the TQRR is a weighted measure of item nonresponse that is the percentage of your total estimate that comes from reported or secondary source data. One drawback to using the TQRR is that the denominator is an estimate, which is not ideal. For the dataset we had, we had to generate a proxy TQRR, as (for one reason) we had to estimate the amount of equivalent quality data.

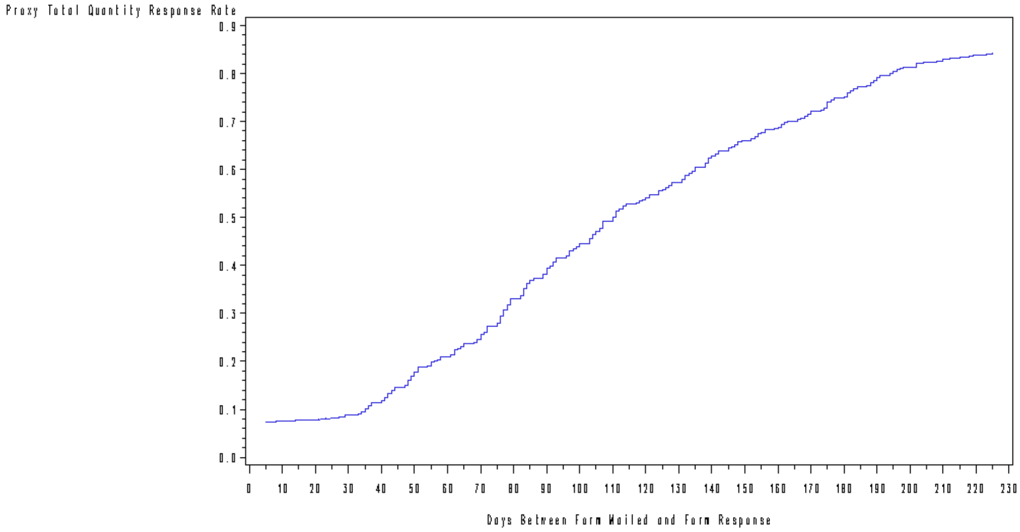


Figure 4. The proxy TQRR for the 2011 ASM.

4.2 R-indicators

Additionally, R-indicators were computed. R-indicators provide a single value between zero and one that measures how close the final sample of respondents are to the sample initially selected. Schouten and Cobben (2007) posit that a response is strongly representative when all individual response probabilities are equal, and weakly representative with respect to some categorical variable *X* when the average response probabilities over the classes of *X* are equal. Strong representativeness is a hypothetical property that cannot be determined in any practical survey setting, as we have no replicates of the response of one single unit. Weak representativeness can be evaluated. As we do not know individual response propensities, we can generate model-based estimates for the individual response propensities and the average response propensity. R-indicators measure deviations from weak representativeness with respect to a vector of available *X*'s. So, as the response propensities become more varied, the R-indicator tends closer to 0. The formula to compute deviance from representativeness is given by:

$$\hat{R}(\hat{\rho}_i) = 1 - 2\hat{S}(\hat{\rho}_i)$$

where

$$\hat{S}(\hat{\rho}_i) = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} d_i(\hat{\rho}_i - \hat{\bar{\rho}})^2}$$

and

N is the number of units in the population,
ρ is the response propensity, and
d_i is the design weight.

As can be seen from the formula, the minimum value of the R-indicator depends on the response rate. For $\bar{p} = 0.5$ it has a minimum value of 0. For $\bar{p} = 0$ or $\bar{p} = 1$, no variation is possible and the minimum value is 1. To model response propensities across measures of size, geography, and types of industry, we developed a weighted logistic regression model using six-digit NAICS, state, and a categorical variable corresponding to the number of employees as predictor variables. As the ASM stratifies on business size and industry we wanted to include these three variables in our model to verify representativeness balanced across the stratification variables. We then computed predicted response propensities for the 51,829 mail cases and calculated the R-indicator. We did this for six "major events" in the data collection period of the survey. The reason for restricting the analysis to only the six events is that the logistic regression model has to be refit each day, and as there are about 225 days during the collection period, this would have proven unduly burdensome. Thus, the six major events for which we calculated representativeness were when the forms were due, the four nonresponse follow-up waves on the day of mail-out, and the R-indicator at the end of data collection.

The results of the R-indicator calculations are shown in Figure 5. With a response rate of only 6%, the high R-indicator reflects lack of variation in estimated response propensities since so few cases have responded. What is of concern, is that as response rates increase, representativeness continues to decrease indicating average response propensities over the categories are not equal. It therefore appears that current nonresponse follow-up procedures may not be effectively targeting certain classes of nonrespondents. We would expect representativeness to increase around day 170, because it is at that point that we see MU representativeness accelerating. In the follow-ups, all SUs and MUs in the mail stratum receive forms and letters, with phone targeting of the largest establishments.

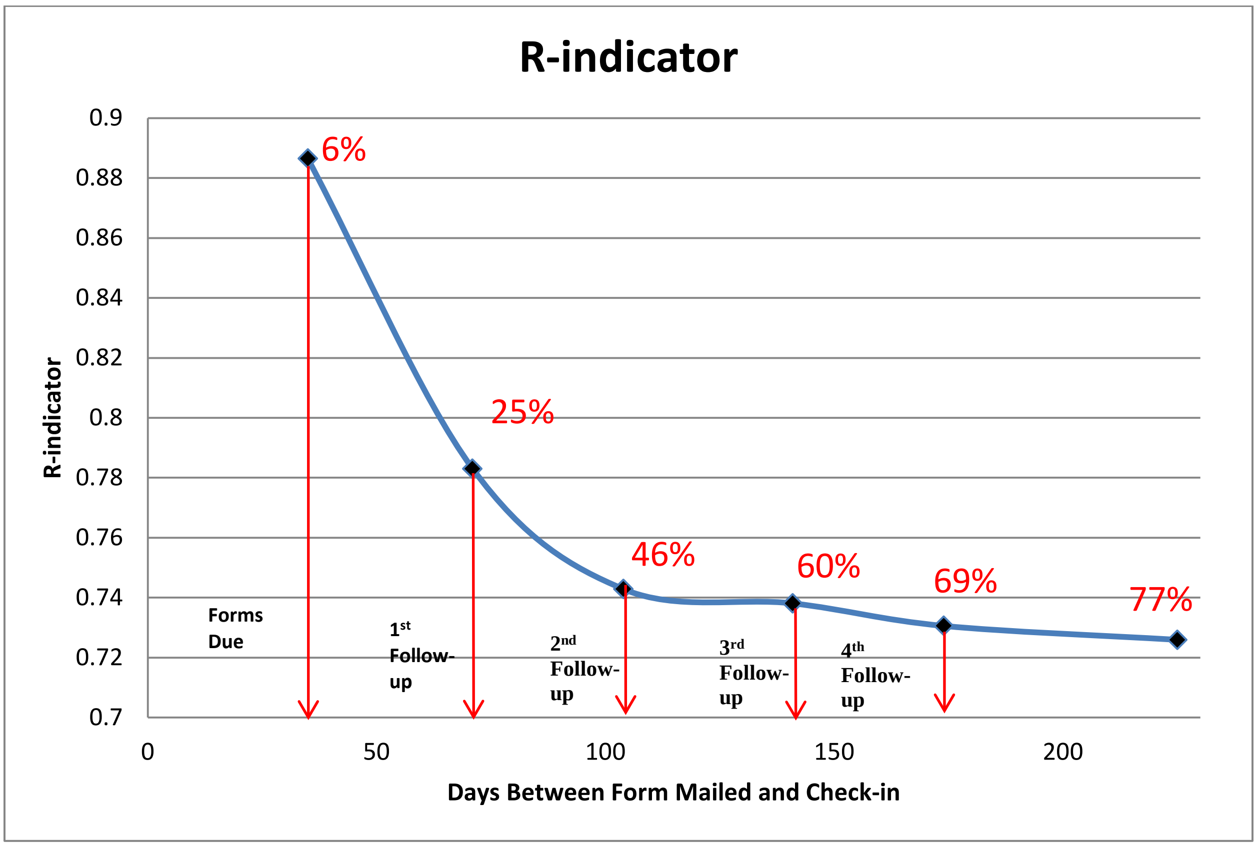


Figure 5. R-indicator calculations for the 2011 ASM on the “due date”, the four nonresponse follow-ups, and when data collection ends. The percentages given in red are the corresponding URRs.

In Figures 6 and 7, the R-indicator is computed where we use 6-digit NAICS or state, respectively, as the only predictor variable in the logistic regression model. We proceeded in this manner to gain insight as to the predictor variable(s) that were most responsible for poor representativeness. As computing the partial R-indicators (see below) computes representativeness estimates across all categories in a categorical variable, this would have become a rather intractable procedure to gain basic insight into what was driving down representativeness, as there were 755 different NAICS categories, 50 state categories, and 24 employment categories.

In Figures 6 and 7 we see the R-indicator beginning to show an increase in the representativeness we expect to see as data collection continues. Nonresponse follow-up is slightly increasing estimated representativeness as reflected by the R-indicator in both industry and geography, but not by much. Because the ASM stratifies by size and industry, and the first R-indicator that builds these two variables into the measure, it is the one that gives a more complete picture of what is going on with how nonresponse follow-up affect representativeness.

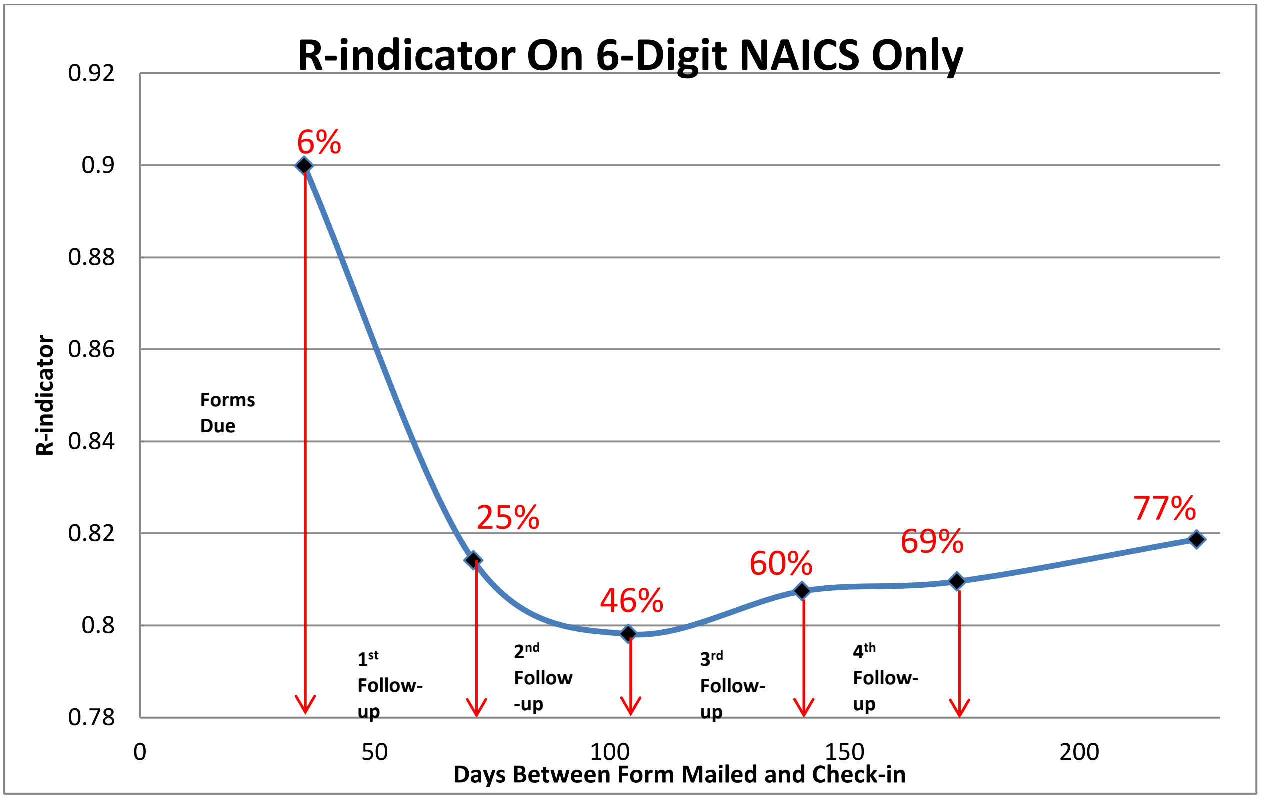


Figure 6. R-indicator calculations for 6-digit NAICS only on the “due date”, the four nonresponse follow-ups, and when data collection ends. The percentages given in red are the corresponding URRs.

In Figure 8, an R-indicator is computed where we use number of employees as the only predictor variable in the logistic regression model. We note that representativeness declines rather sharply. This may be a reflection of follow-up procedures, which concentrate on larger establishments. This decline in representativeness is concerning, as this could be a source of bias since we edit and impute in part using the respondent pool. Larger companies certainly have an impact on confidence interval width for point estimates, but the ASM is a longitudinal survey for which estimates of change are most important. We are thus faced with a tradeoff of trying to get the large establishments versus obtaining a representative sample. This tradeoff becomes more important as we consider the Economic Census, because we consider these results from the ASM as a proxy (in the manufacturing sector) to the Economic Census, to which we benchmark our current surveys.

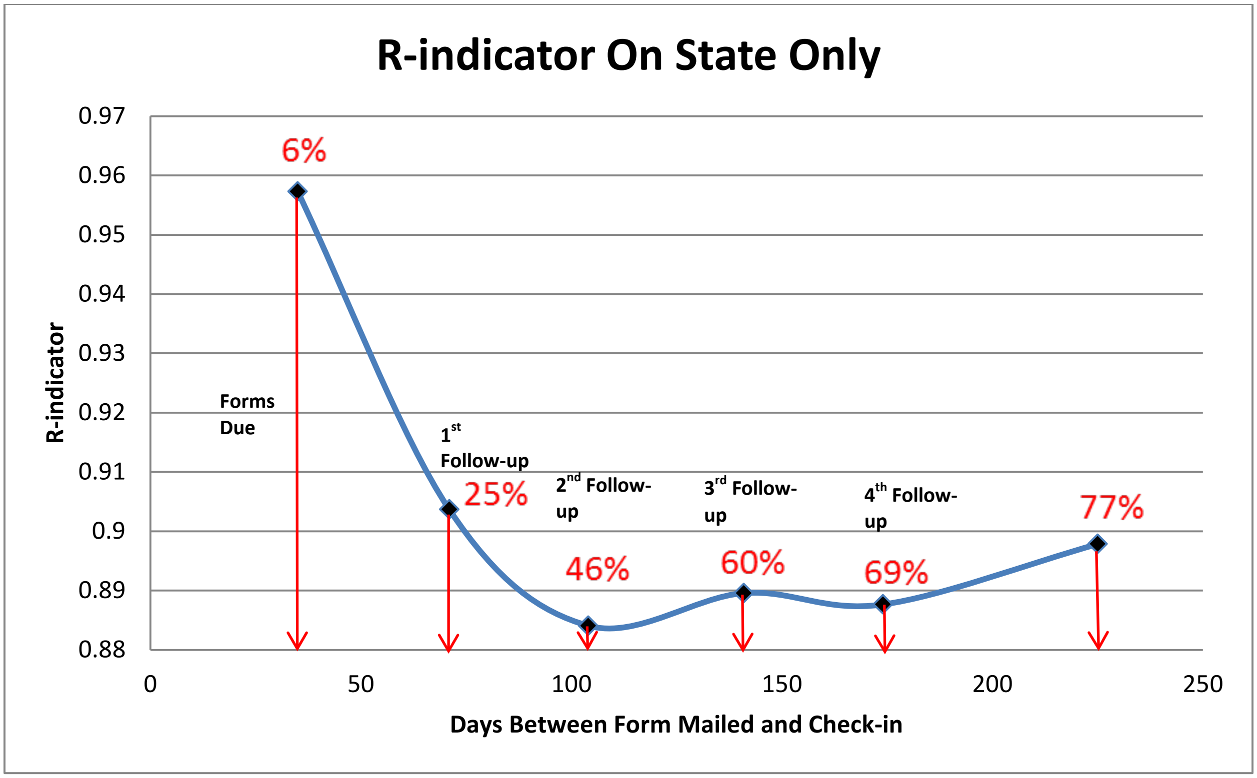


Figure 7. R-indicator calculations for the 2011 ASM for state only on the "due date", the four nonresponse follow-ups, and when data collection ends. The percentages given in red are the corresponding URRs.

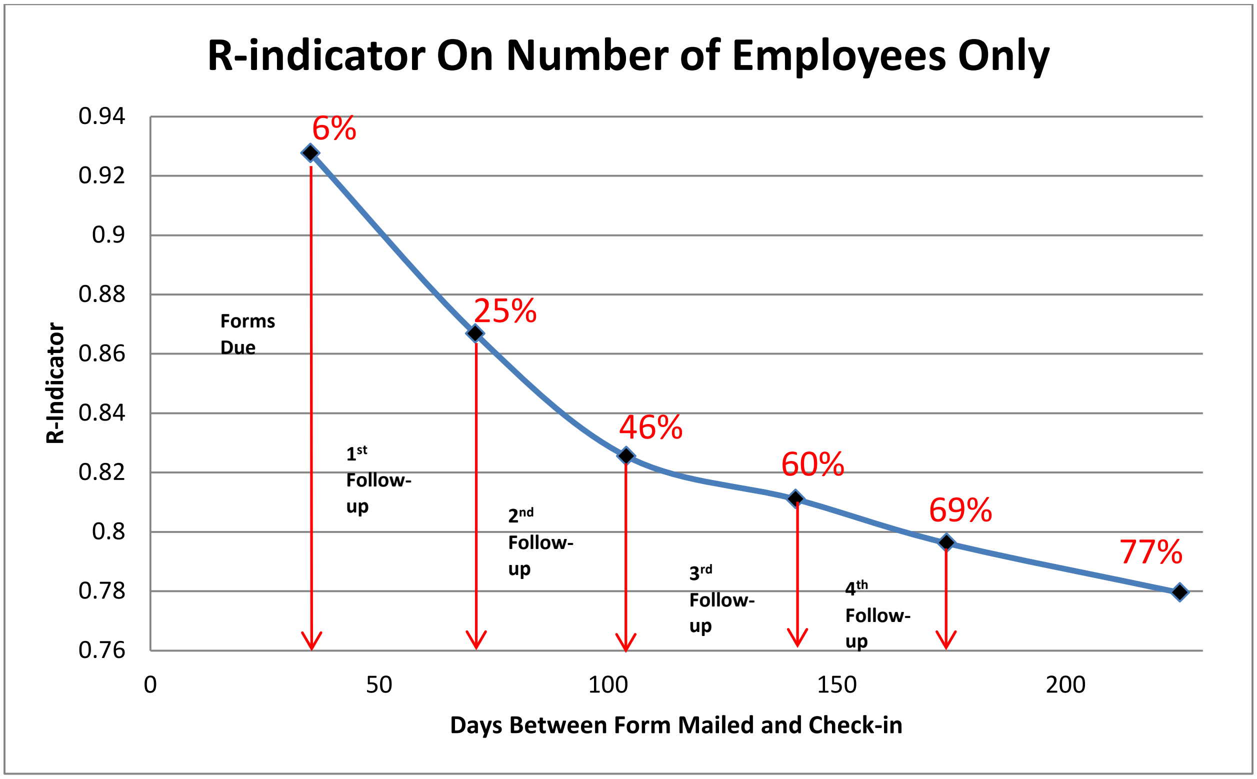


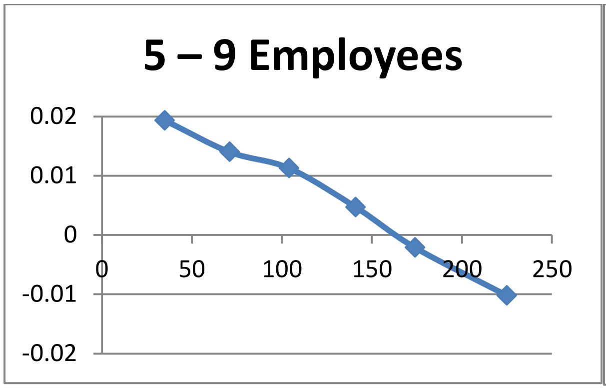
Figure 8. R-indicator calculations for the 2011 ASM for number of employees only on the "due date", the four nonresponse follow-ups, and when data collection ends. The percentages given in red are the corresponding URRs.

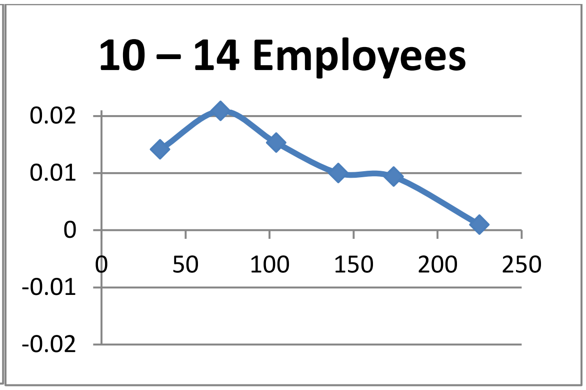
Given the persistent decline in representativeness both overall and in employment, we calculated unconditional partial R-indicators on employment. It should be noted that in constructing the partial R-indicators, we did build all three predictor variables into the model. The employment categories, and their descriptions, are given in Table 1. It should be noted that we collapsed categories after category 87. This was because there were too few businesses in each of the categories to generate a propensity model with accurate estimates. Partial R-indicators are analogous to looking at response rates by key subgroups. The unconditional partial-R indicator, with stratification based on a categorical variable Z with categories $k = 1,2,\ldots,K$ is defined as the variability between categorical response propensities, $S_B(\rho_X|Z)$ where:

$$S_B^2(\rho_X|Z) \cong \sum_{k-1}^{K} \frac{N_k}{N}\left(\bar{\rho}_{X,k} - \bar{\rho}_X\right)^2$$

The average response propensity in stratum k, defined as $\bar{\rho}_{X,k} = \frac{1}{N_k}\sum_{S_k} d_i \rho_X(x_i)$, where $N_k = \sum_{S_k} d_i$. All variables are defined as in the R-indicator formula above, with S_k as the set of sample units in the stratum.

The graphical displays for these partial R-indicators are given below in Figure 9. The overall pattern is quite striking; companies with more employees improve in their representativeness over time relative to companies with fewer employees. This result is congruous with the ASM strategy that targets larger companies more aggressively in the nonresponse follow-up. Again, this is concerning because we edit and impute based on characteristics that are more representative of big companies than of the sample we initially drew, our estimates will most likely be biased. Furthermore, we may be artificially decreasing the variance associated with our estimates, because we are obtaining more of the same type of respondent.





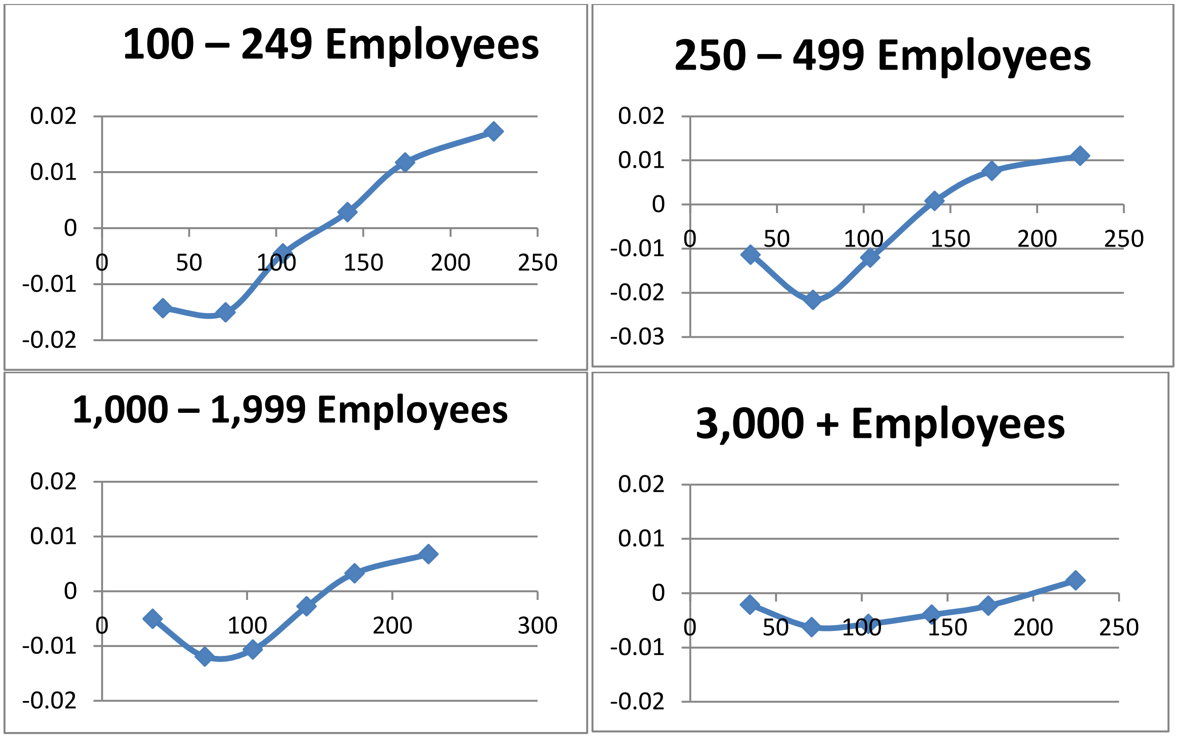


Figure 9. The unconditional partial R-indicators on employment.

5. Discussion

5.1 Research and Economic Survey Programs

As an exploratory research project, the research questions changed considerably from the start of this project until the first draft of this paper because the data differed from what we expected. This research actually stemmed from a paradata project using 2011 ASM electronic paradata. However, as we began building respondent profiles we realized we wanted to get an idea of other metrics aside from response rates that could be used to inform and improve the collection process. In particular, in the Economic Directorate of the Census Bureau we are looking towards real-time metrics to eventually be used in an adaptive design framework. This is a major point because we now know what we would like to have for future paradata/adaptive design research. For instance, we will need a good handle on survey expenses, additional paradata (such as the number and type of error messages, survey break-off information, etc.), and auxiliary data from other business surveys to make this project a success. In a perfect setting, this information is housed in a central location or easily accessible to every survey program. Additionally, there is a need to develop quality measures that update daily. (Such quality measures stem nicely from programs set up for flow processing.) In short, programs can expect up-front costs before we can expect long-term gains.

5.2 Discussion of results

For both the URR and proxy TQRR, we see the rates initially increasing quickly after the due dates, and then slowing over time. Upon examining the URR figures, some follow-ups have a larger percentage increase in associated costs than others, but do not appear to yield any appreciable increase in the number of responses, while other follow-ups have an appreciable increase in the number of responses, but are not associated with large increases in costs. However, without more detailed cost data, we are unsure of the exact relationship between capital expenditure and response. There is the possibility the follow-ups help maintain the observed increase in responses, but only a designed experiment would allow for such conclusions. Furthermore, the pattern seen in the partial-R indicators suggests that representativeness is low for smaller companies. Representativeness of respondents (relative to the initially drawn sample) seems to be decreasing over time. Specifically, in later stages of follow-up we are getting more of the same (larger) establishments and representativeness is low for smaller establishments. Finally, as the Census Bureau makes a push towards near-real-time decision making during the collection process, the R-indicator may be a metric generated as a regular part of data collection.

One idea that had been proposed is that we look at the R-indicators both prior to, and after the editing and imputation process. The reasoning behind this proposal was to see if the final dataset resembled the representativeness of the initially drawn sample. Implementing this idea was problematic for two reasons. First, the frame data we used would be virtually unchanged before and after processing. Therefore, the R-indicator would not change much. Second, if we were to treat all imputed values as respondents, the R-indicator would take on a value of one, simply because there would be no variation in the response propensities.

The outstanding question becomes, given cost and quality indicators, could we be just as effective for less cost? Further investigation is warranted.

5.3 Future research

This research is only scratching the surface of using paradata to examine business-reporting patterns. We are continuing to incorporate the cost information into our analysis; most notably, we hope to be able to show how resources are being allocated throughout the survey life cycle in hopes that we can find ways to improve efficiency. As mentioned above, even if no interventions seem obvious, we may be able to research ways via experimental design to shorten the data collection period, increase conversion of paper respondents to electronic respondents, or test the effects of altering follow-up procedures. Thus, no matter where the research takes us, paradata will prove to be an indispensable tool to create effective models, allowing us to save costs while maintaining high quality in the survey estimates. As we continue to add cost and quality indicators, we will get a better idea how to approach the ASM in an adaptive design context.

Acknowledgements: We wish to thank Eric Merriman for supplying us with Business Register data. We thank Robert Struble for his review of this paper and help in answering questions on the ASM. We thank Michelle Vile Karlsson and Michael Zabelsky for help in supplying us with cost data. We thank Jen Beck, Xijian Liu, Stephen Kaputa for reviewing previous versions of this paper. Finally, we thank Diane Willimack and Stephanie Coffey for reviewing this paper.

References:

Anderson, A., Morrison, R. L., and Brady, C. F. 2005. “Applying Knowledge of Business Survey Response Processes to the Design of Data Collection Software at the U.S. Census Bureau.” Paper prepared for presentation at the *2005 Federal Committee on Statistical Methodology Conference*, Washington, D.C.

Couper M. P. 1998. “Measuring Survey Quality in a CASIC Environment.” *Proceedings of the Section on Survey Research Methods of the American Statistical Association*. Dallas, TX.

de Leew E., de Heer W. 2002. Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. In *Survey Nonresponse*, Groves, RM, Dillman DA, Eltinge JL, Little RJA (eds). Wiley: New York, 41-54.

Fink, E. B., Lineback, J. F., and Hoeffel, E. 2013. “What Can Paradata Tell Us About Business Reporting?” Paper prepared for presentation at the *American Association of Public Opinion Research 68th Annual Conference*, Boston, MA.

Groves, R. M. 2006. Nonresponse rates and nonresponse bias in household surveys. Public Opinion Quarterly, 70, 646-675.

Groves, R. M., Heeringa S. G. 2006. Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs. *Journal of the Royal Statistical Society*, 169, 439-457.

Hak, T., Anderson, A. E., and Willimack, D. K. 2003. Determinants of Web Reporting: A Qualitative Study of Mode Selection. Paper prepared for presentation at the *Federal Committee on Statistical Methodology*, Washington, D.C.

Rao S. W., Glickman M. E., and Glynn R. J. 2008. Stopping Rules for Surveys with Multiple Waves of Nonrespondent Follow-up. *Statistics in Medicine*, 27, 2196-2213.

United States Census Bureau. 2012. U.S. Census Bureau Statistical Quality Standards. Located at: http://www.census.gov/quality/standards/Quality_Standards.pdf.