

NONRESPONSE RATES AND NONRESPONSE BIAS IN HOUSEHOLD SURVEYS

ROBERT M. GROVES

Abstract Many surveys of the U.S. household population are experiencing higher refusal rates. Nonresponse can, but need not, induce nonresponse bias in survey estimates. Recent empirical findings illustrate cases when the linkage between nonresponse rates and nonresponse biases is absent. Despite this, professional standards continue to urge high response rates. Statistical expressions of nonresponse bias can be translated into causal models to guide hypotheses about when nonresponse causes bias. Alternative designs to measure nonresponse bias exist, providing different but incomplete information about the nature of the bias. A synthesis of research studies estimating nonresponse bias shows the bias often present. A logical question at this moment in history is what advantage probability sample surveys have if they suffer from high nonresponse rates. Since postsurvey adjustment for nonresponse requires auxiliary variables, the answer depends on the nature of the design and the quality of the auxiliary variables.

Introduction

One unique value of sample surveys as a tool to document human thought and behavior is their ability to describe large populations without bias and within measurable levels of uncertainty. While the mathematical probability theories required for this inference are a century old (Pearson 1903), and their application to samples of humans about 70 years old (Neyman 1934), recently we have been reminded that this power of surveys is dependent on full measurement

ROBERT M. GROVES is professor and director, University of Michigan Survey Research Center, and research professor at the Joint Program in Survey Methodology. Support from the National Science Foundation (SES-0207435) is gratefully acknowledged. Ashley Bowers, Sungwoo Lim, and Emilia Peytcheva assisted in the compilation of the prior studies on nonresponse bias. The author appreciates comments on an earlier draft from Norman Bradburn, Michael Brick, Mick Couper, Cheryl Eavey, Hermann Habermann, Richard Lempert, Kristen Olson, Emilia Peytcheva, Stanley Presser, Trivellore Raghunathan, Eleanor Singer, Roger Tourangeau, Richard Valliant, and anonymous reviewers. Errors remaining and views expressed are the author's. Address correspondence to the author; e-mail: bgroves@isr.umich.edu.

doi:10.1093/poq/nfl033

© The Author 2006. Published by Oxford University Press on behalf of the American Association for Public Opinion Research. All rights reserved. For permissions, please e-mail: journals.permissions@oxfordjournals.org.

of a probability sample. That is, the original theories assume that nonresponse is absent. However, in the past few decades, developed countries have seen an increase in the rate of sample persons *not* being measured (de Leeuw and de Heer 2002). Hence, it is important to understand the potential impact of nonresponse on the ability of surveys to describe large populations.

Recent articles suggest that changes in nonresponse rates do not necessarily alter survey estimates (Curtin, Presser, and Singer 2000; Keeter et al. 2000; Merkle and Edelman 2002). However, the most common prescription for survey researchers is to minimize nonresponse rates. For example, Alreck and Settle (1995, p. 184) say, "It's obviously important to do as much as possible to reduce nonresponse and encourage an adequate response rate." Babbie (2007, p. 262) is bold enough to say, "A review of the published social research literature suggests that a response rate of at least 50 percent is considered adequate for analysis and reporting. A response of 60 percent is good; a response rate of 70 percent is very good." Finally, Singleton and Straits (2005, p. 145) note, "Therefore, it is very important to pay attention to response rates. For interview surveys, a response rate of 85 percent is minimally adequate; below 70 percent there is a serious chance of bias." All of these quotations come from books used to teach students about survey methods.

This combination of observations—an inferential paradigm that requires 100 percent response rates, declining response rates, evidence that nonresponse rates do not predict nonresponse bias, and rules of thumb that urge survey practitioners to maximize response rates—together seem a recipe for confusion among practitioners.

For these reasons, now is a useful time to synthesize the literature on nonresponse rates and nonresponse bias in surveys. This article (1) reviews statistical notions of nonresponse bias,¹ (2) evaluates different designs for assessing nonresponse biases, (3) reviews the research literature on the relation of nonresponse rates and nonresponse bias, and (4) updates the discussion on the relative merits of probability sampling in the presence of nonresponse. It emphasizes studies of persons (versus organizations), one-time surveys (versus longitudinal surveys), and surveys in the United States and Western Europe (versus the developing world).

What Do We Know about the Linkage between Nonresponse Rates and Nonresponse Bias?

The expressions for nonresponse bias in survey estimates come in various forms. Early survey researchers used approaches that assumed nonresponse was a fixed

1. "Nonresponse error" is often used to refer both to nonresponse bias (i.e., departures of the expected value of an estimate from its true value) and nonresponse error variance (i.e., variations over replications of a survey implementation in the departure of the realized estimate from its expected value due to nonresponse). Because most attention of the field concerns nonresponse bias, we will focus on that type of error in the article.

property of an individual (see Särndal and Lundström 2005, pp. 1–3, for a discussion). What results from that perspective is that the bias of nonresponse in a respondent mean (e.g., mean number of doctor visits in the last 6 months) could be expressed as:

$$\text{Bias}(\bar{y}_r) = \left(\frac{M}{N} \right) (\bar{Y}_r - \bar{Y}_m)$$

where

$\text{Bias}(\bar{y}_r)$ = the nonresponse bias of the unadjusted respondent mean;

\bar{y}_r = the unadjusted mean of the respondents in a sample of the target population;

\bar{Y}_r = the mean of the respondents in the target population;

\bar{Y}_m = the mean of the nonrespondents in the target population;

M = the number of nonrespondents in the target population; and

N = the total number in the target population

so that the respondent mean differs from the mean of the full target population by a function of the nonresponse rate and the difference between respondent and nonrespondent means.² Note that this expression implicitly assumes all other sources of bias (especially measurement error) are absent.

In a sense, this expression is compatible with the notion of a hard-core non respondent group. But with evidence that response rates vary greatly over surveys of different designs of the same population, starting in the 1980s more researchers became attracted to the view that everyone is potentially a respondent or a nonrespondent, depending on circumstances (Lessler and Kalsbeek 1992). That is, everyone has an unobservable “propensity” (a probability, a likelihood) of being a respondent or a nonrespondent, which can be represented by ρ_i . With this viewpoint another expression, one more helpful at the design stage of a survey, approximates the bias of the respondent mean (Bethlehem, 2002):

$$\text{Bias}(\bar{y}_r) \approx \frac{\sigma_{y\rho}}{\bar{\rho}}$$

2. The expression contains population quantities. It is useful to note that a sample-based version of it, $\frac{m}{n} (\bar{y}_r - \bar{y}_m)$, does not have an expected value equal to the population expression, but rather includes a term involving the covariance between the nonresponse rate, on one hand, and the difference between respondent and nonrespondent means, on the other. This covariance term measures whether the distinctiveness of nonrespondents (relative to respondents) changes as the nonresponse rate changes.

where σ_{yp} = the population covariance between the survey variable, y , and the response propensity, p , and \bar{p} is the mean propensity in the target population over sample realizations, given the sample design, and recruitment realizations, given a recruitment protocol design.³

The expression notes that the likelihood of responding is a random variable, varying over conceptual replications of a specified recruitment protocol, and that nonresponse bias is a function of how correlated the survey variable is to the propensity to be measured in the target population. As with the first expression, it is assumed that measurement errors are absent.

In short, nonresponse bias occurs as a function of how correlated response propensity is to the attributes the researcher is measuring. Within the same survey, different sample estimates can be subject to different nonresponse biases. Some, unrelated to the propensity to respond, can be immune from biasing effects of nonresponse; others, in the same survey, can be subject to large biases.

Other types of estimates have their own expressions for nonresponse bias. For example, differences of subclass means (e.g., differences between male and female mean number of doctor visits) would have a simple extension of the above:

$$\text{Bias } [(\bar{y}_{r1} - \bar{y}_{r2})] \approx \frac{\sigma_{y_1 p_1}}{\bar{p}_1} - \frac{\sigma_{y_2 p_2}}{\bar{p}_2}$$

where the subscripts 1 and 2 index the two subclasses. It is this expression that leads to the hope that nonresponse biases might cancel across subclass means. This expression also shows that such a fortuitous result is by no means necessary.

To summarize, what can we deduce from these expressions that contain both nonresponse rates and nonresponse bias?

- a. Nonresponse bias can vary across different statistics in the same survey; thus, low response rate surveys are not necessarily “bad” per se, but they may yield some statistics subject to large nonresponse bias.
- b. Decreasing nonresponse rates may not always lead to lower nonresponse bias; some ways of decreasing nonresponse may increase σ_{yp} .
- c. Components of nonresponse are likely to differentially affect σ_{yp} for a given y . Noncontact propensities are likely to be correlated with different y 's than refusal propensities or noninterview propensities due to health, language, etc.
- d. To discern when nonresponse rates portend nonresponse bias, we must understand how the influences for and against participation are related

3. As we note later in this article, \bar{p} , the propensity to respond, is often the result of a multistep process of accessing (contacting) the sample unit, delivering the survey request, and obtaining cooperation from the sample unit, each step of which might be viewed as a stochastic process, conditional on the outcome of the prior step.

- to the survey measures. This is likely to vary across ways of obtaining higher response rates, modes of data collection, population studied, etc.
- e. There is no minimum response rate below which survey estimates are necessarily subject to bias.
 - f. An important area of scientific inquiry is whether certain recruitment protocols themselves alter the covariance between y and p as the response rate changes (e.g., incentives reducing the covariance, as in Groves, Singer, and Corning [2000]).

THINKING CAUSALLY ABOUT NONRESPONSE BIAS

The expressions above for the stochastic view of nonresponse bias of the sample mean are a function of covariances between response propensities and the survey variable. Indeed, if survey designs could somehow produce recruitment protocols that resulted in zero covariances for all survey variables, nonresponse biases could be eliminated for sample means. However, many advances in science arise from theory-building about the causes of correlations among variables. Hence, it may be useful to reformulate the problem in terms of causal models that describe alternative conditions related to nonresponse bias. In doing so, we move from an operational level to a conceptual level. Further, we suspend attention to how these causal assertions might be operationalized or tested.

Figure 1 shows five possible situations relevant to a covariance between y and p in a survey recruitment protocol. The first, the “separate causes model,” produces no covariance between p and y . This is the case where there is a set of causes, Z , of response propensity that is distinct and uncorrelated with the causes of the survey variable. In this case there is no nonresponse bias involving y , regardless of the response rate of the survey. Because completely uncorrelated causes (Z and X) are difficult to imagine, this first model is a simplified case but would correspond to notions of “missing completely at random” (Little and Rubin 2002).

The second model, the “common cause model,” generates a covariance between the two attributes because of a common cause of both of them. For example, the construct of “topic interest” could underlie some results of Messonnier, Bergstrom, Cornwell, Teasley, and Cordell’s (2000) survey of users of a recreational lake. Response rates for users whose homes are on the lake are much higher than those of users who live elsewhere. Why? One hypothesis is that they possess a set of interests concerning that particular lake, which are partially met by participating in a survey about that lake. Their interests are causal factors in their survey participation decision. Those same interests can influence some survey variables (e.g., their willingness to pay to maintain the lake quality). This causal model is consistent with notions of “missing at random” (Little and Rubin 2002); that is, within classes homogeneous on interests concerning the lake, there would be no nonresponse bias on estimates

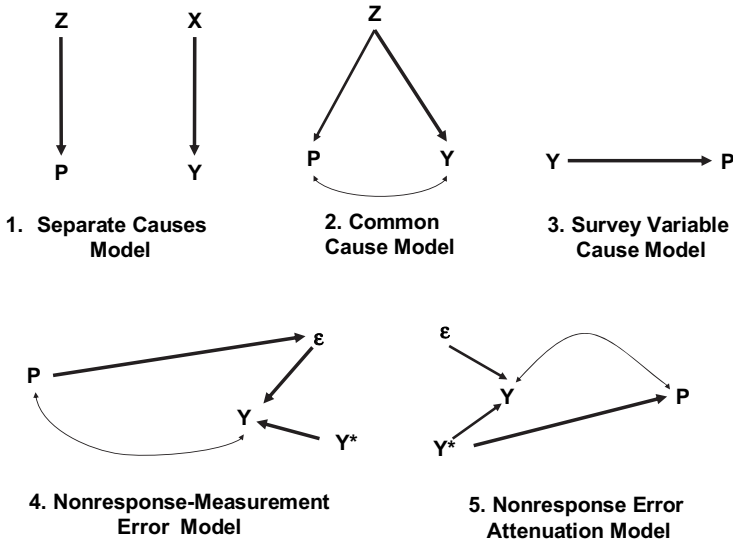


Figure 1. Five idealized causal models of response propensity (P), the reported survey variable (Y), the true value of the survey variable (Y^*) and other variables (X, Z) having different implications for nonresponse bias of the unadjusted respondent mean on the reported survey variable (Y).

of sample means for some survey variables. The graphic for the model implies a single cause, Z ; in practice, there is likely to be a whole set of Z variables. If they are measured in the survey (and on the frame or target population), however, there is hope of statistically eliminating their effects on survey estimates involving Y , through postsurvey adjustment.

The third model, the “survey variable cause model” in figure 1, produces a covariance between p and y through a direct causal relation. That is, the variable of interest is itself the cause of the response propensity. An example of this might be time-use surveys, where a statistic of interest might be the percentage of time spent at home. Respondents’ time spent away from home is a cause of interviewers’ failure to contact a household, given some fixed callback pattern. This leads to overestimates of the proportion of time spent at home in surveys without callbacks. Another obvious (even humorous, in a survey kind of way!) example might be measuring the proportion of the illiterate using a written self-administered questionnaire, where illiteracy is one cause of nonresponse. This third model corresponds to the “not missing at random” case, leading to nonignorability conditions (Little and Rubin 2002).

The fourth model, the “nonresponse-measurement error model,” describes a nexus between nonresponse bias and measurement error. In this model the

level of response propensity determines the magnitude of a measurement error, ϵ , associated with the survey variable, y . Under this model, the survey report is $y_i = y_i^* + \epsilon_i$, a true value (y^*) plus an error term. In this case the expected value of ϵ is nonzero, some over- or under-reporting phenomenon. Because the error term is caused by the propensity, p , there is a covariance between the y and the p . One example of this might be Cannell and Fowler's (1963) finding that respondents interviewed early in the survey period (and assumed to be more motivated to perform the respondent task) produce better reports on their hospital visits than those recruited with more effort later in the survey period. One mechanism that would produce this finding is that of the fourth causal model. That is, the sample mean is biased as a function of response rate because of measurement error differences between easily reached, cooperative respondents and inaccessible, reluctant respondents. (Note: this case is not treated in the nonresponse bias expressions above.) While this is outside the scope of simpler nonresponse processes, it resembles the common cause model's production of a missing at random mechanism. That is, if the measurement error quantities were known, nonresponse bias in estimates based on y could be eliminated.

The fifth model in figure 1, the "nonresponse error attenuation model," shows a situation with no nonresponse bias despite clear links with the survey variable. This occurs with a simple model of measurement error variance, $y_i = y_i^* + \epsilon_i$, where y^* is the true value and y is the reported value, and as is traditionally assumed, there is no covariance between the true value and the response deviation, ϵ_i . This model is most appropriate for survey responses subject to low reliability or high response variance. In this situation any covariance between the response propensity and the y^* is attenuated by the measurement error variance. More simply stated, with this causal structure survey estimates based on very noisy measurements have lower likelihood of nonresponse bias. The fifth model can be viewed as a more general case of the third model (the "survey variable cause" model), one with diminished "not missing at random" features because of measurement error. That is, some of the examples in the research literature of no linkage between survey estimates and nonresponse rates may merely reflect large random measurement errors.

To summarize figure 1, there are three contrasting features of the causal models. First, they differ in the nature of the causal relationships involving p and y that are producing the covariance between p and y . Often these will also affect the magnitude of the covariance, with the second model typically having smaller covariances (the correlation of p and y is the product of two correlations, ρ_{yp} and ρ_{y^*y}). Second, the last two models incorporate possibilities of measurement errors affecting estimated nonresponse biases. Third, the models pose very different challenges for reducing nonresponse bias in postsurvey adjustment.

POSTSURVEY ADJUSTMENT TO REDUCE NONRESPONSE BIAS

The expressions above concern sample values and estimates, unadjusted in any way to compensate for nonresponse. It is common, however, to use weighting class adjustments (Bethlehem 2002), raking (Deville, Särndal, and Sautory 1993), calibration methods (Deville and Särndal 1992; Lundström and Särndal 1999), or propensity models (Ekholm and Laaksonen 1991) to reduce the biasing effects of response propensities correlated with the survey variables. In addition, poststratification, using population totals for subclasses also measured on the respondents, is used both to reduce standard errors and correct for coverage and nonresponse biases (Kalton 1981; Valliant 1993).

All of these adjustment techniques require assumptions that groups of respondents and nonrespondents share response propensities *and* distributional properties on survey measures. For example, when persons within a weighting class are shown to have the same expected values of a survey variable as nonrespondents, the weighting class adjusted mean can eliminate the nonresponse bias of the unadjusted mean.

In practice, the assumptions underlying the adjustment procedures are generally untestable (e.g., whether respondents and nonrespondents within a weighting class have the same values on key variables is unknowable). Further, some adjustments may make matters worse (Brick, Le, and West 2003; Little 1982; Little et al. 1997; Little and Vartivarian 2003).

The nonresponse adjustment procedures can be easily mapped onto the causal models of figure 1. All of them attempt to use Z variables (auxiliary to the substantive purposes of the survey but measured in the survey) to adjust statistics based on Y , to remove the biasing effects of response propensities that are a function of Z . This is completely effective under model 2, the common cause model. In the common cause model, if Z is measured on both respondents and nonrespondents, the researcher can remove the nonresponse bias by weighting class adjustments using the Z variable. (This is equivalent to observing that, controlling for Z , there is no covariance between p and y , or $\rho_{py|Z} = 0$.)

For models other than the common cause model, postsurvey adjustments face greater challenges. If the Z variable is used for adjustment (say, in a weighting class adjustment) in the first model, the separate causes model, there will be no change in estimated means involving Y , but, because Z is unrelated to Y , the standard error of the adjusted estimate would increase. Since there are no possible auxiliary variables for adjustment in model 3, the survey variable cause model, other adjustment models reflecting “nonignorable” mechanisms must be constructed (Rubin 1987), generally requiring much more heroic assumptions. Finally, the traditional nonresponse adjustment techniques do not address the last two models of figure 1, although some structural equation modeling approaches can incorporate them (e.g., Brownstone, Golob, and Kazimi 2002).

Methods for Assessing Nonresponse Bias

As nonresponse rates increase in household surveys, nonresponse bias studies become increasingly important (indeed, they are called for by recent OMB guidelines for U.S. federal government-funded surveys; see Office of Management and Budget [2006]). A review of the nonresponse bias literature allows us to classify the various types of research designs used to assess nonresponse bias. This section reviews the design alternatives, their strengths, and their weaknesses.

RESPONSE RATE COMPARISONS ACROSS SUBGROUPS

We begin with this method because it is easy to perform, even though it does not yield direct estimates of nonresponse bias on key statistics (see, for example, Brick et al. 2003). It can, however, be used to compare respondent and nonrespondent distributions on the subgroup variables. The researcher usually presents estimates of response rates on key subgroups of the target population (e.g., age, race, gender, urbanicity subgroups). Generally, the researcher asserts that there is no evidence of “nonresponse bias” if the response rates are similar across subgroups. If there are low response rate groups, the researcher either argues that they are unimportant for his/her purposes or attempts post-survey adjustment.

Asserting that constant response rates over subgroups imply no nonresponse bias is in essence asserting that the subgrouping variables are the only possible “common causes” of response propensity and survey variables. This is generally an untenable assumption. Hence, the method is one of the least informative about possible nonresponse biases in estimates based on other survey variables.

USING RICH SAMPLING FRAME DATA OR SUPPLEMENTAL MATCHED DATA

Some studies match each person in the sample with individual records from some external database. Others use sampling frames to identify the target population that record many attributes for each population member, so-called rich sampling frames. Using variables on the external data set, the researcher compares respondent and nonrespondent values. This is common in health research studies in which medical records are available for matching. When the record base is used from the start as the sampling frame, no matching is required. Prominent examples of this technique are Kennickell and McManus (1993), Bolstein (1991), Assael and Keon (1982), and Lin and Schaeffer (1995).

The strength of this design is that identical measurements are available for all members of the sample, both respondents and nonrespondents. Thus,

accurate estimates of nonresponse bias *on those frame or external data variables* can be constructed. Further, statistical relationships between those variables and survey variables can be measured among respondents, to address in a partial way the likely nonresponse biases of survey variables.

The weaknesses of the method are that the variables available are, by definition, not all those of key interest to the survey; the record data may be subject to missing values; and there may be measurement error in the record data that damages the nonresponse bias estimates. When measurement errors are suspected, models predicting the survey variable from the record variable, estimated on respondents, may be helpful as an alternative estimate of the survey variable among nonrespondents (e.g., David et al. 1986).

COMPARISONS TO SIMILAR ESTIMATES FROM OTHER SOURCES

Perhaps the most common tool for nonresponse bias analysis is comparing the respondent-based estimates with those from another, more accurate source. An example of this approach in household surveys is to compare the distributions of age, gender, race, and other sociodemographic variables among respondents with those from the most recent census data for the population. Another example is to use as the gold standard a high-quality government survey (e.g., in the United States the Current Population Survey is often used).

The strengths of this method are that estimates independent of the survey in question are compared. When the comparison survey has great credibility among users, then obtaining similar estimates gives some confidence about the survey in question.

The weaknesses of this tool are that the key survey variables of the study do not usually exist in the external source; that the form of the measurements may differ between the focal survey and the gold standard survey (thus, measurement error differences contaminate the comparison); and that the coverage and nonresponse characteristics of the gold standard survey are not completely known.

STUDYING VARIATION WITHIN THE EXISTING SURVEY: NONRESPONSE FOLLOW-UP STUDIES

This technique involves subsetting respondents into subgroups that may exhibit different nonresponse bias characteristics. Examples of this include comparing respondent estimates from early cooperators with those from the full respondent data set (as in Curtin, Presser, and Singer 2000, 2005; Dunkelberg and Day 1973; Lin and Schaeffer 1995), comparing respondents in the first phase sample with those in both the first and second phase samples (as in Groves and Wissoker 1999), and comparing observations made during data collection on both respondents and nonrespondents (Groves and Couper 1998).

The strength of this method is that it can be used in many different modes of data collection, with diverse populations, on diverse topics. The requisite data are process data recording number of call attempts, follow-up mailings, and contacts with the sample unit, etc. All estimates that can be computed from the survey are candidates for this analysis.

The weakness of the method is that it offers no direct information about the nonrespondents to the survey. Instead, the notion of a “continuum of resistance” is often asserted by the analysts, noting that nonrespondents should be most similar to those respondents measured only after great effort was expended. Often, there is no evidence that respondents interviewed only with great effort are different enough and prevalent enough to produce large changes on the estimates examined (Curtin, Presser, and Singer 2000). Other studies have shown that nonrespondents are very different from those measured with great effort (Lin and Schaeffer 1995). Thus, although this is an easy analysis to do if the researcher has access to the process data, it often provides little information about the nonresponse bias that remains after all efforts in the survey have been expended.

The method does appear to be useful for another purpose—asking how reducing the amount of effort expended to obtain interviews would affect the data. Such a question is common on repeated cross-section surveys.

CONTRASTING ALTERNATIVE POSTSURVEY ADJUSTMENTS FOR NONRESPONSE

This class of nonresponse bias studies attempts to measure the amount of nonresponse bias that might be eliminated by postsurvey adjustment. Examples of this are comparisons of unadjusted respondent-based estimates with estimates utilizing some weighting class adjustments (Ekholm and Laaksonen 1991; Izazola-Licea et al. 2000; Potthoff, Manton, and Woodbury 1993), comparing nonresponse-adjusted estimates with those also including poststratification, comparing unadjusted respondent-based estimates with fully imputed data set estimates, and various combinations of these methods.

The strength of this technique is that a large set of alternative estimators purporting to measure the same population parameter can be compared. When the alternative estimators are based on very different assumptions about the nature of the nonresponse *and* are similar in magnitude, the researcher can have more confidence in the conclusions from the survey. If they differ, the researcher has some reason for caution.

The weakness of such methods is that they are limited to observing *differences* in nonresponse bias associated with alternative estimates. They lack an unambiguous gold standard because each of the adjustment schemes requires some untestable assumptions. If, for example, different adjustments yield different estimates, the analyst has no guidance on which is preferable, lacking some external benchmark.

SUMMARY OF NONRESPONSE BIAS STUDY DESIGNS

Each of the techniques has value. Some (like the use of frame variables) may involve shifts in mode of data collection or an increase in coverage error in exchange for estimation of nonresponse bias; others are relatively cheap (like comparisons to census data) but limited in information about the nonresponse bias of key survey estimates. All of them can be characterized as attempting to measure directly or indirectly the covariance between response propensity and the survey variables.

As response rates decline, researchers face a growing obligation to mount nonresponse bias studies in order to inform the evaluation of survey estimates. Because of the diverse properties of the techniques above, it is wise to study nonresponse biases using multiple methods simultaneously.

When Do Nonresponse Rates Portend Nonresponse Bias?

The recent studies of Keeter et al. (2000), Curtin, Presser, and Singer (2000), and Merkle and Edelman (2002) lead to the impression that nonresponse rates are a much smaller threat to survey estimates than suggested by prior practical guidance. However, the articles need to be placed in the context of years of methodological research. In the extreme, they are misinterpreted as implying that there is rarely, if ever, a reason to worry about nonresponse bias.

This section of the article summarizes results from a large number of research studies that present estimates of nonresponse bias (see the appendix for the list of references). The articles result from a search of a wide variety of electronic databases for literature on survey nonresponse, including the Scholarly Journal Archive (JSTOR), Gale/Info Trac Expanded Academic ASAP, ABI/INFORM Global, LexisNexis, Proquest Research Library, SilverPlatter databases, OCLC Social Science Abstracts, ECO and ArticleFirst databases, SocioFile, ISI Web of Knowledge, Web of Science, Social Sciences Citation Index and ISI Proceedings, and ScienceDirect. Searches of journals with a specific focus on survey methodology, such as *Public Opinion Quarterly* and *Journal of Official Statistics*, and searches of survey methodology reference books, such as *Nonresponse in Household Interview Surveys* (Groves and Couper 1998), were also performed. *Proceedings of the American Statistical Association Survey Research Methods Section* and papers presented at the 1999 International Conference on Survey Nonresponse were reviewed. In addition, general Google Internet searches for survey nonresponse literature were conducted, as well as specific searches for nonresponse studies from the Survey of Consumer Finances and National Center for Education Statistics surveys. Then references to other work cited in these articles were pursued.

Associated with this article is material on the *Public Opinion Quarterly* Web site that provides compact presentations of the research studies. For each study, there is a short description of the survey design, a table or graph showing

evidence of the link between nonresponse rates and nonresponse biases, and a slide with conclusions. A majority of the 30 articles were published in medical journals. The method of data collection for half of the studies was mail surveys. Five studies used face-to-face interviewing, and five others were telephone surveys; the remainder used diverse mixes of modes.

All of the studies report estimated means (or percentages) without postsurvey adjustments. There are 235 separate estimates from the 30 articles. While the documentation is not uniform across the articles, the vast majority report a response rate that most resembles the AAPOR response rate 1 (AAPOR 2006), eliminating ineligible sample persons from the denominator. The mean nonresponse rate over the 235 estimates is 35 percent (median = 30 percent); there are few surveys with very low nonresponse rates (i.e., less than 15 percent). Thus, it appears that few formal studies of nonresponse bias are mounted for surveys with low nonresponse rates.

We present several different figures based on the 30 articles, each of which plots a slightly different function of nonresponse effects. All figures share various features—the x -axis is the nonresponse rate of the survey; each point corresponds to an unadjusted respondent mean on some variable in the study; estimates from the same study thus form a vertical line, permitting the reader to see variation among different variables *within* studies.

Some adjustments to the raw nonresponse bias estimates reported in the articles are desirable to improve the comparability among them. First, the bias estimates are computed on diverse units of measurement of the individual statistics. Some are dollar amounts (because the estimate is mean dollars); some are percentages, where the magnitudes of the biases are heavily influenced by the size of the percentage. Second, the individual points are means based on very different sample designs and sizes. That is, they are subject to different sampling errors. Finally, the variables measured are not comparable across surveys. It is possible that the studies with low nonresponse rates measure variables that have lower or higher correlations with response propensities than those with high nonresponse rates.

Figure 2 standardizes the units of measurement, removing the sign of the bias by taking absolute values and including complementary values for binary variables as separate estimates. This yields 335 estimates from 30 articles. This is one attempt to equalize the units of measurement, but it does not control for differences in sampling variance of the estimates. The figure shows a plot of the absolute values of percentage relative nonresponse bias:

$$\left| \frac{100 \cdot (\bar{y}_r - \bar{y}_n)}{\bar{y}_n} \right|$$

The mean percentage absolute relative bias is 8.7 percent of the full sample estimate. The correlation between the nonresponse rate and the percentage

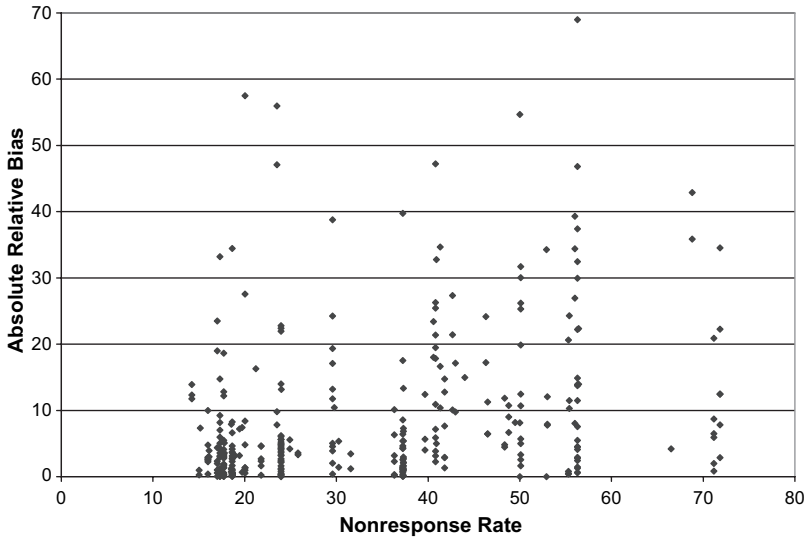


Figure 2. Percentage absolute relative nonresponse bias of $\left| \frac{100 * (\bar{y}_r - \bar{y}_n)}{\bar{y}_n} \right|$ 235 respondent means by nonresponse rate from 30 different methodological studies.

absolute nonresponse bias is 0.33, a modest positive correlation; the square of the correlation is 0.11. The vast majority of the variation in relative nonresponse bias lies *within* surveys, not *among* them.⁴

Another useful transformation would be to compute the nonresponse bias on standardized variables,

$$y_i^{std} = \frac{(y_i - \bar{y}_n)}{\sigma_y}$$

where y_i^{std} is the standardized value of the y variable on the i th sample unit, and σ_y is the standard deviation of the y variable. This standardizes both the units of measurement and the standard deviation of the estimates (but does not adjust for differential standard errors of the estimates). Unfortunately, the published articles do not in general provide standard deviations of the y variables.

4. With a larger number of studies, a useful approach would be a formal meta-analysis in an attempt to identify what properties of individual estimates make them susceptible to nonresponse bias. Ideally, the data would permit a study of the relationship of the survey measurements to their sensitivity to nonresponse bias.

However, 191 of the 235 estimates are percentages. On binary variables producing percentage estimates it is possible to compute an estimate of σ_y via $\sqrt{\bar{y}_n(100 - \bar{y}_n)}$, approximately unbiased. The nonresponse bias, $\bar{y}_r^{std} - \bar{y}_n^{std}$, is presented in figure 3 for standardized binary variables. The mean bias for the 191 estimates is -0.00031 , quite near zero; the mean absolute bias is 0.054 , or 5.4 percent of the standard deviation of the y variable. The correlation between the nonresponse rate and the absolute value of the estimated bias is 0.36 ; its square is 0.13 . As with figure 2, much of the variation in nonresponse bias lies within studies, not across studies differing in nonresponse rates.

Finally, in figure 4 we address the differential sampling errors of the estimates. Another adjustment to the raw numbers estimates the difference between the respondent and the nonrespondent estimates, $\bar{y}_r - \bar{y}_m$, removing the effects of differential sampling errors. We note that $E(\bar{y}_r^{std} - \bar{y}_m^{std})^2 = \{E(\bar{y}_r^{std} - \bar{y}_m^{std})\}^2 + Var(\bar{y}_r^{std} - \bar{y}_m^{std})$, where $Var(\bar{y}_r^{std} - \bar{y}_m^{std})$ is the variance of the difference of the two means and \bar{y}_r^{std} and \bar{y}_m^{std} . Thus,

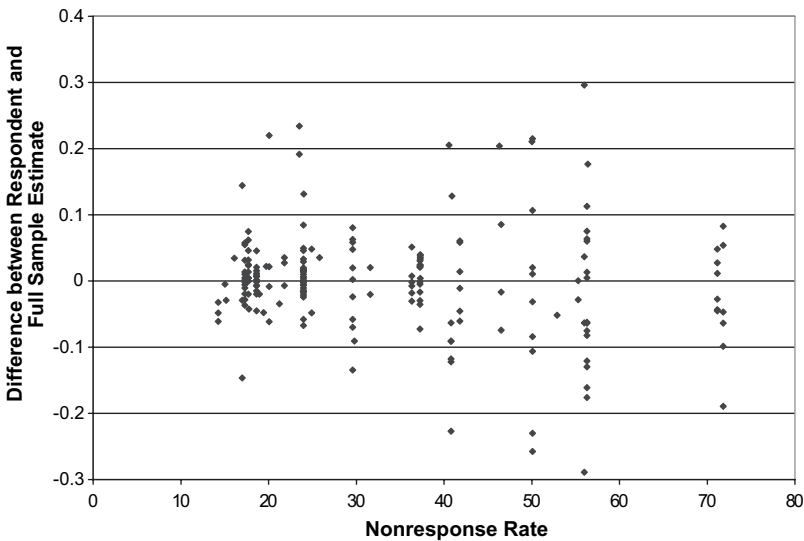


Figure 3. Estimated nonresponse bias for standardized values ($\bar{y}_r^{std} - \bar{y}_n^{std}$) of 191 percentages by nonresponse rate from 23 different methodological studies.

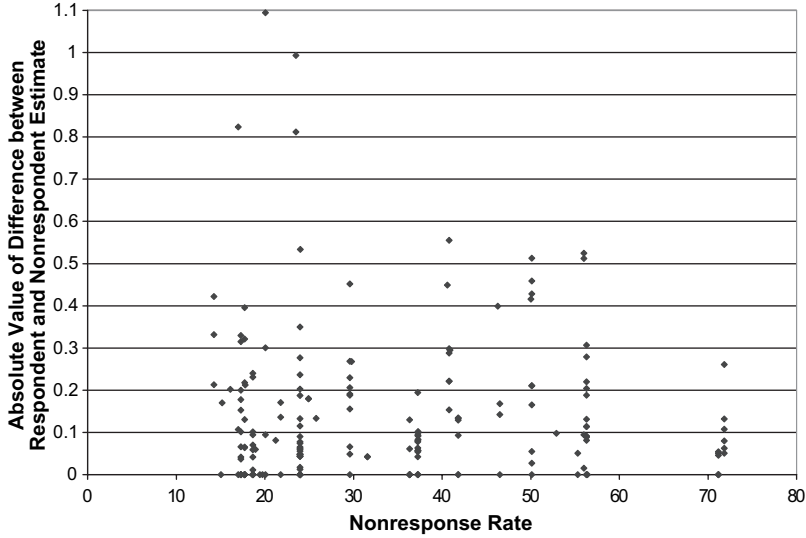


Figure 4. Estimated absolute difference between respondent and nonrespondent percentages for standardized variables $|(\bar{y}_r^{std} - \bar{y}_m^{std})|$, adjusted for sampling variance, for 191 percentages by nonresponse rate from 23 different methodological studies.

$E[(\bar{y}_r^{std} - \bar{y}_m^{std})^2 - \text{var}(\bar{y}_r^{std}) - \text{var}(\bar{y}_m^{std})] \approx \{E(\bar{y}_r^{std} - \bar{y}_m^{std})\}^2$, where $\text{var}(\bar{y}_r^{std})$ and $\text{var}(\bar{y}_m^{std})$ are sample-based estimates of variance.⁵ To ease comparisons with the other figures, figure 4's y-axis is the square root of the quantity above, estimating $|\bar{y}_r^{std} - \bar{y}_m^{std}|$ for standardized percentages (the same set as that in figure 3). Using the expression $\text{Bias}(\bar{y}_r) = \left(\frac{M}{N}\right)(\bar{Y}_r - \bar{Y}_m)$, the y-axis in figure 4 thus represents a sample-based estimate of the absolute value of the difference term. If this term were a constant over all estimates for a given survey, then nonresponse bias would be a simple function of response rate. Figure 4 attempts to remove all the artifacts plaguing the comparison of estimates, except for the inherent variability in the correlation between the measures and response propensity (i.e., the σ_{yp}).

5. In practice, especially with small sample sizes, the estimates of $\{E(\bar{y}_r^{std} - \bar{y}_m^{std})\}^2$ can be negative. These were set to zero values in figure 4.

Figure 4 has a different shape for its plotted values. In contrast to figures 2 and 3, visually it appears that there is greater variability in $|\bar{y}_r^{sd} - \bar{y}_m^{sd}|$ for studies with lower rather than higher nonresponse rates. This follows from the fact that bias is the product of the y-axis value and the proportion of nonrespondents, deflating variation in bias for low nonresponse rates. Over all 23 articles, the mean value of the estimated difference is 0.14, which might be thought of as 14 percent of a standard deviation of the y variable.⁶ As expected, the correlation between the nonresponse rate and the respondent-nonrespondent differences is very low, -0.017 ; its square is 0.00028 . The chief finding from figure 4, as from the others, is that much variation in bias lies *within* surveys rather than *among* surveys.

We are now ready to draw overall conclusions from figures 2–4. The first conclusion is that there is ample evidence that nonresponse bias does occur. The mean percentage relative nonresponse bias for the 335 respondent estimates is about 9 percent (median = 4 percent); for the 191 respondent percentages the average bias is about 5 percent of a standard deviation (median = 3 percent). One of the highest relative nonresponse biases in figure 2 (55 percent) is for a contingent valuation study regarding the quality of lake water. The estimate of interest is the percentage of lake users who are lake residents, with the percentage of lake residents among respondents greatly overestimating the percentage of lake users in the total sample (Messonnier et al. 2000). Other high levels of relative nonresponse bias (36 percent and 43 percent) apply to estimates of mean wealth and mean income, based on use of tax records, among a very wealthy population (Kennickell and McManus 1993). In this study, respondent wealth greatly underestimates wealth levels in the full sample because the wealthiest tend not to respond. In short, it is clear from the figures that nonresponse bias happens.

The second conclusion from figures 2–4 is that, while nonresponse bias clearly does occur, the nonresponse rate of a survey alone is not a very good predictor of the magnitude of the bias. The percentage of variation in the bias indicators “explained” by nonresponse rates is very low in each of figures 2–4, despite their using different functions of nonresponse bias. The correlations are low because they do not reflect characteristics of respondent estimates that make them sensitive to nonresponse bias. The relationships between response propensities and the y variables measured in the various studies are themselves highly variable within studies. In short, nonresponse rate *alone* is a weak predictor of nonresponse bias components.

6. To check the effects of sampling variability, we computed the estimated average nonresponse bias from figure 4 (multiplying the response rate of the study by $[(\bar{y}_r^{sd} - \bar{y}_m^{sd})]$) as 0.047 versus 0.054 using $(\bar{y}_r^{sd} - \bar{y}_n^{sd})$ in figure 3. Thus, both forms of estimation yield average nonresponse biases that are about 5 percent of a standard deviation.

The third observation about figures 2–4 is really a caution against misinterpretation. Using the figures, for example, it is not appropriate to make comments about the effect of a change in response rates on bias *within* a survey. For example, one of the points in figure 2 represents relative nonresponse biases of about 40 percent on a wealth index. These are estimates from a survey with a 68 percent nonresponse rate. What would happen to nonresponse bias if the nonresponse rate were lowered, perhaps through incentives, refusal conversion, and so forth, from 68 percent to 63 percent? Figure 2 does not answer that question. If the nonresponse rate were reduced by methods more attractive to the higher-income persons, then the relative nonresponse bias might decrease dramatically. If the nonresponse rate were reduced by methods equally attractive to higher- and lower-income persons, then the bias might be reduced as a simple function of how much the nonresponse rate declined. However, if the nonresponse rate were reduced by methods more attractive to the lower-income persons, then the nonresponse bias might actually increase, despite a lower nonresponse rate. Figure 2 does not give us any information about which of the three outcomes is likely.

The fourth conclusion from figures 2–4 is that they alone do not identify the circumstances under which nonresponse rates are related to nonresponse bias. The figures are based on a collection of studies that vary on mode, topic, use of incentives, target population, measurement techniques, and a host of other factors that could themselves affect the relationship between the overall nonresponse rate and the relative nonresponse bias. The figures merely show that, *net of all those design features*, there is no strong relationship between a survey's nonresponse rate and the nonresponse biases of its diverse estimates. The challenge to the field is to build and test appropriate theories about mechanisms that link response propensities to nonresponse biases. The figures make it clear that the theory must be articulated at the level of the individual measure, not at the level of a survey.

Let's synthesize the observations thus far. First, statistical expressions for nonresponse bias make it clear that only the *risk* of nonresponse bias (not nonresponse bias itself) is reduced with decreasing nonresponse rates. Said differently, higher response rates do not necessarily reduce nonresponse bias for any survey or any given estimate. If lower nonresponse rates are obtained by attracting people with unusual values on the survey variables, the σ_{yp} term might become larger in absolute value; if lower nonresponse rates are obtained by devices equally attractive to all remaining nonrespondents, then σ_{yp} might move toward 0.0.

Second, if we examine in a meta-analytic way what the survey methodological literature finds for the linkage between nonresponse rates and nonresponse biases, we find large nonresponse biases for some statistics but no strong empirical relationship between response rates and nonresponse bias.

Third, this allows us to deduce that, in practice, for any given nonresponse rate, nonresponse biases should be expected to vary across estimates within

the same survey. The biases are heavily influenced by the covariance between response propensities and the particular survey variables. The covariance itself is a statistical manifestation of causal systems underlying the survey participation decision *and* causal systems of the survey variables measured. In short, nonresponse bias is a phenomenon much more complex than mere non-response rates.

Nonresponse and the Survey Practitioner

UBIQUITOUS MAIN EFFECTS ON RESPONSE PROPENSITY

There are a few attributes that appear to be predictive of response propensities in a wide variety of survey settings. Some of the findings are compatible with a common cause model in figure 1. For example, sponsorship effects are common, with central government surveys generating higher response rates than academic surveys, and academic surveys generating higher response rates than commercial surveys (Groves and Couper 1998). Further, positive or negative affect toward the sponsor of the survey may be related to the survey variables measured (e.g., as when central government surveys ask about receipt of welfare benefits) and to response propensity. (If affect toward the sponsor is measured directly, then the "survey variable cause model" might apply.) This probably explains why customer satisfaction surveys are rarely conducted directly by the service provider.

Burden of the survey, as measured by pages in a self-administered questionnaire, produces lower response rates (Goyder 1985; Heberlein and Baumgartner 1978); burden as measured by length of telephone and face-to-face interviews shows less clear effects (Bogen 1996).

Males refuse more than females (Smith 1983). Urbanicity is a powerful indicator of response rates in all modes (de Leeuw and de Heer 2002). Adults who live alone tend to be refusals (Groves and Couper 1998); households with young children show higher response rates than others (Lievesley 1988).

Many other attributes (e.g., socioeconomic status, racial minority status) appear to have variable effects over modes, topics, or sponsors of surveys.

In at least some surveys, these influences on survey participation are correlated with the variables of interest in the survey. The practitioner must decide whether this is likely to be the case and whether, therefore, differential effort should be assigned to the groups with low base propensities. To assign more effort to subgroups with low base propensities requires identifying them. Rich sampling frames are sometimes useful for this; collecting observations as part of the survey process (e.g., single-person households) is also useful. If response rates are increased using devices that are not disproportionately attractive to the low propensity groups, then nonresponse biases may increase despite lowered nonresponse rates.

DEVICES FOR INCREASING RESPONSE PROPENSITIES THAT POTENTIALLY AFFECT NONRESPONSE BIASES

The past two decades have clearly shown that survey researchers have the power to alter the response propensities of the persons they sample. The modern survey researcher is equipped with a large set of tools that act to raise the response propensity of persons exposed to them. While much of the literature has focused on whether a tool raises the overall response rate to a survey, the discussion above shows that nonresponse bias is a function of the covariance between response propensity and variables involved in a survey estimate. Hence, a rereading of the response rate literature is required to know whether a response rate increase is good or bad in a particular survey. For example, there can be increases in nonresponse bias with increasing response rates when persons with distinctive values on the survey variable are differentially sensitive to the design feature creating higher response propensities.

Such a rereading is not easy because the research literature on advance letters, incentives, interviewer persuasion, mode effects, and the like has focused much more on response rates than nonresponse bias. This section provides a brief glance at the more readily available literature on the differential effects of design features on response propensities. That is, the review identifies statistical interaction effects in experimental studies on response rates.

Advance Letters. Advance letters are routine in mail and face-to-face surveys (Dillman 1978; Luppés 2000). They have been found to increase response rates, on average, in household surveys (de Leeuw et al. 2005). If the letters explicitly note the sensitive content of the interview, however, they can depress response rates (as in ACSF Group 1992, cited in de Leeuw et al. 2005). Letters on market research firm stationery depressed response rates, but university stationery increased response rates (Brunner and Carroll 1969). Advance letters in random digit dial (RDD) surveys increase response rates only for cases with mailable addresses.

Can advance letters affect nonresponse bias? Unfortunately, there are few studies showing differential effects of letters across subgroups. One would speculate that nonresponse biases sensitive to advance letters would be located among correlates of the likelihood of receiving or reading the letters. For example, since advance letters require literacy for some of their effects, one would expect letters to have lower effects in semiliterate populations.

Incentives. Incentives have been shown to be effective in increasing overall response rates in all modes of surveys (Singer 2002). Several empirical studies show that incentives bring into the respondent pool sample persons who are uninterested in the survey topic. Roberts, Roberts, Sibbald, and Torgerson (2000), in a survey about hormone replacement therapy, show 8.5 percentage points fewer respondents who have had the therapy with incentives than without; Baumgartner, Rathbun, Boyle, Welsh, and Laughland (1998), in a study about time-of-day utility pricing, show that using incentives leads to a higher

percentage of respondents who had opted out of the pricing program; Groves, Presser, and Dipko (2004) show lower percentages of persons with statuses likely to be interested in the survey topic in the incentive group than in the group without incentives. These interaction effects are consistent with the leverage-salience theory of survey participation (Groves, Singer, and Corning 2000), which argues that in the absence of intrinsic motives for survey participation (e.g., self-interest in discussing a particular topic), incentives are extrinsic substitutes. Nonresponse biases related to incentive effects are likely to vary among estimates correlated with socioeconomic status and levels of interest in the survey topic.

Perhaps the most dramatic example of potentially harmful effects of increasing response rates is the incentive experiment reported by Merkle, Edelman, Dykeman, and Brogan (1998). In this exit poll experiment, a pen incentive increased overall response rates. However, the incentive increased Democratic Party voters' response propensities more than those of Republicans. As a result, the higher response rate condition (with incentives) had larger nonresponse bias for vote statistics than the lower response rate condition (without incentives).

Interviewer Workloads and Callback Rules. It is well accepted that repeated callbacks are effective at reducing noncontact nonresponse (Goyder 1985; Heberlein and Baumgartner 1978) and that large interviewer workloads can reduce the ability of interviewers to make such callbacks (Botman and Thornberry 1992). Thus, reduced callback surveys disproportionately produce nonresponse among those less frequently accessible to interviewers. For example, Hilgard and Payne (1944) show underestimates of employment rates and average number of children per household among those interviewed on the first call. Groves, Wissoker, Greene, McNeeley, and Montemarano (2001) show that persons who live by themselves are disproportionately missed in a survey using few callbacks. Nonresponse biases related to callbacks are likely in estimates related to time use.

Observable Interviewer Attributes. It is a common practice to attempt to assign interviewers to cases that have similar age, race, ethnicity, or other observable attributes. The empirical support for this is stronger for measurement errors (Schuman and Converse 1971) than for nonresponse biases (see Brehm 1994). There is evidence that female interviewers obtain higher participation rates from female respondents than male interviewers do (Nealon 1983). If survey variables were strong correlates of gender (e.g., purchases of dresses), then nonresponse biases might result in estimates involving those variables. Nonresponse biases related to observable attributes of interviewers are likely in variables having similar causes to the respondents' attitudes toward those attributes (e.g., racial attitudes).

This brief review of how design features that increase response rates may affect nonresponse bias is woefully inadequate. Its limitation should act as a call to designers of response rate experimental studies to look for interaction

effects (differences in treatment effects across subgroups). New studies and reanalyses are sorely needed. The review is sufficient, however, to justify the conclusion that survey practitioners should look beyond the mere response rate-enhancing impacts of advance letters, callbacks, incentives, and other common tools. The effect of these tools on nonresponse bias (versus nonresponse rate) depends on whether the tools have differential effects on groups differing on the variables of interest in the survey.

With High Nonresponse Rates, Why Use Probability Sampling?

Many nonprobability sample designs attempt to balance respondents on a set of attributes correlated with the survey variables, thus assuring that respondents resemble population distributions on those variables. This balancing is obtained by quotas or other mechanisms to achieve the targeted number of interviews, with relatively little effort at follow-up or refusal conversion. With probability sampling, both repeated callbacks and refusal conversion are required. But given the rising costs of achieving higher response rates *and* the findings of few nonresponse biases in lower response rate surveys, some in the field are questioning the value of the probability sampling framework for surveys. They ask, “What advantage does a probability sample have for representing a target population if its nonresponse rate is very high and its achieved sample is smaller than that of nonprobability surveys of equal or lower cost?”

A thorough answer to this question, updating the initial discussions of Neyman (1934), Stephan and McCarthy (1958), Stephenson (1979), Smith (1983), Deville (1991), Smith (1994), and others, is needed. Since there is a near-infinite variety of nonprobability sample designs, the question defies a simple answer. It is noncontroversial to note, however, that by departing from randomized selection, those designs burden the analyst with adjusting respondent estimates both for nonrandomized selection procedures and for nonresponse (see Chang and Krosnick 2001).

All statistical adjustments for nonobservation (in both probability and nonprobability samples) require some auxiliary variables measured on the respondents and available for the full population or at least the nonrespondents. Whether the nonprobability sample survey can fulfill the heavier adjustment burdens is a function of what auxiliary variables are available. Nonprobability samples with explicit frames (e.g., address samples, RDD samples using quota schemes to select persons) generally have more auxiliary variables for adjustment than nonprobability designs that merely use volunteer samples (e.g., mall intercept surveys, volunteer Internet polls). Without a frame the adjustment models are subject to more questions, because then adjustment, for example in household surveys, is based on census data (or high-quality probability samples

like the Current Population Survey), and the auxiliary variables in the survey may be measured by a different process than those on the population data.

But just as we know that increasing response rates can increase bias when those brought into the respondent pool are distinctive on the survey variables, so too there is no guarantee that adjustments won't create more bias in the estimates (Brick, Le, and West 2003; Little and Vartivarian 2003). Further, the assumptions of the models are generally untestable, given the data at hand.

Thus, both probability and nonprobability sample surveys require answers to the same set of questions (but they have different resources to answer them): (1) Is the subset of the target population eligible for selection potentially biased on the survey variables? (Surveys with explicit frames can be compared with the target population.) (2) Is the subset actually sampled potentially biased on the survey variables? (Probability samples can answer "no" because of random sampling.) (3) Is the subset measured (among those selected) potentially biased on the survey variables? and (4) Do the adjustment procedures fulfill their assumptions? (Surveys rich in auxiliary variables have an advantage.)

Resources external to the survey itself make a difference in our ability to answer these questions—rich sampling frames and randomized selection provide links to the target population even in the presence of nonresponse. Under some circumstances, probability samples with high nonresponse that are drawn from sparse sampling frames may lose out to nonprobability samples from rich sampling frames with powerful adjustment models. Since probability samples have the advantage of eliminating bias at the selection step, it is useful to consider assembling auxiliary variables to compensate for nonresponse bias through the use of strong postsurvey adjustment models.

Putting It All Together—What's a Survey Practitioner to Do?

The discussion thus far, while reviewing the current state of knowledge about the linkage between nonresponse rates and nonresponse bias, does not itself yield guidance to the practicing survey researcher. This section provides my own practical deductions:

1. *Blind pursuit of high response rates in probability samples is unwise; informed pursuit of high response rates is wise.*

We know that the easiest way to reduce nonresponse rates in most household surveys is to reduce the noncontact rate; the most difficult, to reduce refusals. In most household surveys, increasing response rates through reducing noncontacts merely exacerbates urban-rural disparities in response rates. As another example, in RDD surveys, mailing advance letters to increase response rates merely increases the already higher listed number response rate versus the unlisted number response rate. Whether these imbalances are important depends on the survey variables of interest. Response rate improvement

efforts should be guided by some knowledge of how groups likely to be affected by the efforts relate to key survey variables. Sometimes, increasing the response rate can increase the σ_{yp} term by increasing the propensity to respond of groups distinctive on y (e.g., the attractiveness of the Merkle et al. [1998] pen incentive to Democrats in an exit poll).

2. *Despite low response rates, probability sampling retains the value of unbiased sampling procedures from well-defined sampling frames.*

Coverage error of well-defined sampling frames can be evaluated relative to a desired target population, prior to the survey being launched. Probability sampling of the frame permits use of auxiliary variables on the frame to improve the estimation from the respondent-based data. Volunteer panels lose these advantages. Low response rate probability sample surveys need to marshal the power of auxiliary variables for postsurvey adjustment.

3. *Collecting auxiliary variables on respondents and nonrespondents to guide attempts to balance response rates across key subgroups is wise.*

Informed response rate reduction requires auxiliary variables. The best auxiliary variables are those simultaneously correlated with response propensity and the key survey variables. An important obligation of survey designers is to identify and collect auxiliary variables useful for adjustment purposes.

4. *Examining alternative postsurvey adjustments to repair imbalances remaining after data collection is wise.*

It is clear that efforts to achieve 100 percent response rates are naive in most U.S. household surveys. Every survey should examine alternative postsurvey adjustments as a sensitivity test of the assumptions made by the adjustments.

5. *Using multiple approaches to assess nonresponse bias on key survey estimates is wise.*

Attempting whenever possible to use frames with large numbers of auxiliary variables, supplementing the sample with nonsurvey variables (through interviewer observation, use of commercial data, etc.), double samples of nonrespondents, mode switches, and the like can all provide some insight into the linkages between response propensities and the survey variables. Mounting such studies is a new obligation of the practitioner.

Summary and Conclusions

Nonresponse to household surveys is growing, inflating the costs of surveys that attempt to achieve high response rates. However, nonresponse biases in estimates are only indirectly related to nonresponse rates. A key parameter

determining the nexus between nonresponse rates and nonresponse bias is how strongly correlated the survey variable of interest is with response propensity, the likelihood of responding. Thus, estimates within a survey are likely to vary in their nonresponse biases. However, current rules of thumb of good survey practice dictate striving for a high response rate as an indicator of the quality of all survey estimates.

Assembly of methodological studies whose designs permit estimation of nonresponse bias shows that empirically there is no simple relationship between nonresponse rates and nonresponse biases. That is, research results comport with the assertion that covariances between survey variables and response propensities are highly variable across items within a survey, survey conditions, and populations. Hence, there is little empirical support for the notion that low response rate surveys de facto produce estimates with high nonresponse bias.

Because of falling response rates, legitimate questions are arising anew about the relative advantages of probability sample surveys. Probability sampling offers measurable sampling errors and unbiased estimates when 100 percent response rates are obtained. There is no such guarantee with low response rate surveys. Thus, within the probability sampling paradigm, high response rates are valued. Unfortunately, the alternative research designs for descriptive statistics, most notably volunteer panels, quota samples from large compilations of personal data records, and so forth, require even more heroic assumptions to derive the unbiased survey estimates.

As nonresponse rates increase, however, effective surveys require the designer to anticipate nonresponse and actively seek auxiliary data that can be used to reduce the effect of the covariance of response propensities and the survey variables.

Supplementary Data

Supplementary data are available online at <http://pubopq.oxfordjournals.org/>.

Appendix

REFERENCES FOR NONRESPONSE BIAS STUDIES

- Assael, Henry, and John Keon. 1982. "Nonsampling vs. Sampling Errors in Survey Research." *Journal of Marketing* 46:114–23.
- Barchielli, Alessandro, and Daniela Balzi. 2002. "Nine-Year Follow-Up of a Survey on Smoking Habits in Florence (Italy): Higher Mortality among Non-Responders." *International Journal of Epidemiology* 31:1038–42.
- Bolstein, Richard. 1991. "Comparison of the Likelihood to Vote among Preelection Poll Respondents and Nonrespondents." *Public Opinion Quarterly* 55:648–50.

Cohen, G., and J. C. Duffy. 2002. "Are Nonrespondents to Health Surveys Less Healthy Than Respondents?" *Journal of Official Statistics* 18:13–23.

Criqui, Michael, Elizabeth Barrett-Connor, and Melissa Austin. 1978. "Differences between Respondents and Non-Respondents in a Population-Based Cardiovascular Disease Study." *American Journal of Epidemiology* 108:367–72.

Dallosso, Helen, R. James Matthews, Catherine McGrother, Michael Clarke, S. Perry, Christine Shaw, and Carol Jagger. 2003. "An Investigation into Nonresponse Bias in a Postal Survey on Urinary Symptoms." *British Journal of Urology* 91:631–36.

Drew, James, and Robert Groves. 1989. "Adjusting for Nonresponse in a Telephone Subscriber Survey." In *Proceedings of the American Statistical Association, Survey Research Methods Section*, pp. 452–56. Alexandria, VA: American Statistical Association.

Etter, Jean-Francois, and Thomas Perneger. 1997. "Analysis of Nonresponse Bias in a Mailed Health Survey." *Journal of Clinical Epidemiology* 50(10):1123–28.

Goldberg, Marcel, Jean Francois Chastang, Annette Leclerc, Marie Zins, Sébastien Bonenfant, Isabelle Bugel, Nadine Kaniewski, Annie Schmaus, Isabelle Niedhamer, Michèle Piciotti, Anne Chevalier, Catherine Godard, and Ellen Imbernon. 2001. "Socioeconomic, Demographic, Occupational, and Health Factors' Association with Participation in a Long-Term Epidemiologic Survey: A Prospective Study of the French GAZEL Cohort and Its Target Population." *American Journal of Epidemiology* 154:373–84.

Grosset, Jane. 1994. *The Biasing Effects of Nonresponses on Information Gathered by Mail Surveys*. Institutional Report no. 78. Philadelphia: Community College of Philadelphia.

Hudson, Darren, Lee-Hong Seah, Diane Hite, and Tim Haab. 2004. "Telephone Pre-surveys, Self-Selection, and Non-Response Bias to Mail and Internet Surveys in Economic Research." *Applied Economics Letters* 11:237–40.

Kendrick, Denise, Rhyddian Hapgood, and Patricia Marsh. 2001. "Do Safety Practices Differ between Responders and Non-Responders to a Safety Questionnaire." *Injury Prevention* 7:100–103.

Kennickell, Arthur, and Douglas McManus. 1993. "Sampling for Household Financial Characteristics Using Frame Information on Past Income." In *Proceedings of Survey Research Methods Section of the American Statistical Association*, pp. 88–97. Alexandria, VA: American Statistical Association.

Khare, Meena, Leyla Mohadjer, Trena Ezzati-Rice, and Joseph Waksberg. 1994. "An Evaluation of Nonresponse Bias in NHANES III (1988–91)." In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 949–54. Alexandria, VA: American Statistical Association.

Kim, Jane, Jess Lonner, Charles Nelson, and Paul Lotke. 2004. "Response Bias: Effect on Outcomes Evaluation by Mail Survey after Total Knee Arthroplasty." *Journal of Bone and Joint Surgery* 86A(1):15–21.

Lahaut, Vivienne, Harrie Jansen, Dike van de Mheen, and Henk Garretsen. 2002. "Non-Response Bias in a Sample Survey on Alcohol Consumption." *Alcohol and Alcoholism* 3:256–60.

Lin, I-Fen, and Nora Schaeffer. 1995. "Using Survey Participants to Estimate the Impact of Nonparticipation." *Public Opinion Quarterly* 59:236–58.

McNutt, Louise-Anne, and Robin Lee. 2000. "Intimate Partner Violence Prevalence Estimation Using Telephone Surveys: Understanding the Effect of Nonresponse Bias." *American Journal of Epidemiology* 152:438–41.

Melton, L. J., D. Dyke, J. Karnes, and Peter O'Brien. 1993. "Nonresponse Bias in Studies of Diabetic Complications: The Rochester Diabetic Neuropathy Study." *Journal of Clinical Epidemiology* 46:341–48.

Messonnier, Mark, John Bergstrom, Christopher Cornwell, R. Jeff Teasley, and H. Ken Cordell. 2000. "Survey Response-Related Biases in Contingent Valuation: Concepts, Remedies, and Empirical Application to Valuing Aquatic Plant Management." *American Journal of Agricultural Economics* 83:438–50.

Paganini-Hill, Annlia, Grace Hsu, A. Chao, and Ronald Ross. 1993. "Comparison of Early and Late Respondents to a Postal Health Survey Questionnaire." *Epidemiology* 4:375–79.

Pedersen, Peder. 2002. "Non-Response Bias: A Study Using Matched Survey-Register Labour Market Data." Working Paper no. 02-02. Aarhus, Denmark: Centre for Labour Market and Social Research.

Perneger, Thomas, Eric Chamot, and Patrick Bovier. 2005. "Nonresponse Bias in a Survey of Patient Perceptions of Hospital Care." *Medical Care* 43:374–80.

Potter, D. E. B. 1989. "Nonresponse in a Survey of Nursing Home Residents." In *Proceedings of the American Statistical Association, Survey Research Methods Section*, pp. 440–45. Alexandria, VA: American Statistical Association.

Reigneveld, Sijmen, and Karien Stronks. 1999. "The Impact of Response Bias on Estimates of Health Care Utilization in a Metropolitan Area: The Use of Administrative Data." *International Journal of Epidemiology* 28:1134–40.

Rogelberg, Steven, Alexandra Luong, Matthew Sederburg, and Dean Cristol. 2000. "Employee Attitude Surveys: Examining the Attitudes of Noncompliant Employees." *Journal of Applied Psychology* 85:284–93.

Sheikh, K., and S. Mattingly. 1981. "Investigating Non-Response Bias in Mail Surveys." *Journal of Epidemiology and Community Health* 35:293–96.

Teitler, Julien, Nancy Reichman, and Susan Sprachman. 2003. "Costs and Benefits of Improving Response Rates for a Hard-to-Reach Population." *Public Opinion Quarterly* 67:126–38.

van Kenhove, Patrick, Katrien Wijne, and Kristof de Wulf. 2002. "The Influence of Topic Involvement on Mail-Survey Response Behavior." *Psychology and Marketing* 19:293–301.

Voogt, Robert, and Hetty van Kempen. 2002. "Nonresponse Bias and Stimulus Effects in the Dutch National Election Study." *Quality and Quantity* 36:325–45.

Walsh, John, Sara Kiesler, Lee Sproull, and Bradford Hesse. 1992. "Self-Selected and Randomly Selected Respondents in a Computer Network Survey." *Public Opinion Quarterly* 56:241–44.

References

- AAPOR, 2006. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Survey*. 4th Edition. Lenexa, Kansas: AAPOR.
- ACSF Group. 1992. "Analysis of Sexual Behavior in France (ACSF): What Kind of Advance Letter Increases the Acceptance Rate in a Telephone Survey Behavior." *Bulletin de Methodologie Sociologique* 35:46–54.
- Alreck, Pamela, and Robert Settle. 1995. *The Survey Research Handbook*. New York: McGraw-Hill.

- American Association for Public Opinion Research (AAPOR). 2005. "Standards and Best Practices." Available online at <http://www.aapor.org/standards.asp> (accessed October 25, 2006).
- Assael, Henry, and John Keon. 1982. "Nonsampling vs. Sampling Errors in Survey Research." *Journal of Marketing* 46:114–23.
- Babbie, Earl. 2007. *The Practice of Social Research*. 11th ed. Belmont, CA: Wadsworth.
- Baumgartner, Robert, Patricia Rathbun, Keven Boyle, Michael Welsh, and Drew Laughland. 1998. "The Effect of Prepaid Monetary Incentives on Mail Survey Response Rates and Response Quality." Paper presented at the annual meeting of the American Association for Public Opinion Research, St. Louis, MO.
- Bethlehem, Jelke. 2002. "Weighting Nonresponse Adjustments Based on Auxiliary Information." In *Survey Nonresponse*, ed. Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, pp. 275–88. New York: Wiley.
- Bogen, Karen. 1996. "The Effect of Questionnaire Length on Response Rates: A Review of the Literature." In *Proceedings of the Section on Survey Research Methods*, pp. 1020–25. Alexandria, VA: American Statistical Association.
- Bolstein, Richard. 1991. "Comparison of the Likelihood to Vote among Preelection Poll Respondents and Nonrespondents." *Public Opinion Quarterly* 55:648–50.
- Botman, Steven, and Owen Thornberry. 1992. "Survey Design Features Correlates of Nonresponse." *Proceedings of the Survey Research Methods Section*, pp. 309–14. Alexandria, VA: American Statistical Association.
- Brehm, John. 1994. *The Phantom Respondents: Opinion Surveys and Political Representation*. Ann Arbor: University of Michigan Press.
- Brick, J. Michael, David Ferraro, Teresa Strickler, and Carin Rauch. 2003. *No. 8: 2002 NSAF Response Rates*. Washington, DC: Urban Institute. Available online at http://www.urban.org/UploadedPDF/900692_2002_Methodology_8.pdf (accessed October 25, 2006).
- Brick, J. Michael, Thanh Le, and Jerry West. 2003. "Dealing with Movers in a Longitudinal Study of Children." In *Proceedings of Statistics Canada Symposium 2003*, pp. 2–12. Ottawa: Statistics Canada.
- Brownstone, David, Thomas Golob, and Camilla Kazimi. 2002. "Modeling Nonignorable Attrition and Measurement Error in Panel Surveys: An Application to Travel Demand Modeling." In *Survey Nonresponse*, ed. Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, pp. 373–87. New York: Wiley.
- Brunner, Allen, and Stephen Carroll. 1969. "The Effect of Prior Notification on the Refusal Rate in Fixed Address Surveys." *Journal of Advertising* 9:42–44.
- Cannell, Charles, and Floyd Fowler. 1963. *A Study of the Reporting Visits to Doctors in the National Health Survey*. Ann Arbor, MI: Survey Research Center.
- Chang, Lin, and Jon Krosnick. 2001. "The Representativeness of National Samples: Comparisons of an RDD Telephone Survey with Matched Internet Surveys by Harris Interactive and Knowledge Networks." Paper presented at the annual meeting of the American Association for Public Opinion Research, Montreal.
- Curtin, Richard, Stanley Presser, and Eleanor Singer. 2000. "The Effects of Response Rate Changes on the Index of Consumer Sentiment." *Public Opinion Quarterly* 64:413–28.
- . 2005. "Changes in Telephone Survey Nonresponse over the Past Quarter Century." *Public Opinion Quarterly* 69:87–98.
- David, Martin, Roderick Little, Michael Samuhel, and Robert Triest. 1986. "Alternative Methods of CPS Income Imputation." *Journal of the American Statistical Association* 81(393):29–41.
- de Leeuw, Edith, and Wim de Heer. 2002. "Trends in Household Survey Nonresponse: A Longitudinal and International Comparison." In *Survey Nonresponse*, ed. Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, pp. 41–54. New York: Wiley.
- de Leeuw, Edith, Joop Hox, Elli Korendijk, and Gerty Lensvelt-Mulders. 2005. "The Influence of Advance Letters on Response in Telephone Surveys: A Meta-Analysis." Unpublished manuscript.
- Déville, Jean-Claude. 1991. "A Theory of Quota Surveys." *Survey Methodology* 17:163–81.
- Déville, Jean-Claude, and Carl Särndal. 1992. "Calibration Estimators in Survey Sampling." *Journal of the American Statistical Association* 87:367–82.
- Déville, Jean-Claude, Carl Särndal, and Oliver Sautory. 1993. "Generalized Raking Procedures in Survey Sampling." *Journal of the American Statistical Association* 88:1013–20.
- Dillman, Don. 1978. *Mail and Telephone Surveys: The Total Design Method*. New York: Wiley.

- Dunkelberg, William, and George Day. 1973. "Nonresponse Bias and Callbacks in Sample Surveys." *Journal of Marketing Research* 10:160–68.
- Ekholm, Anders, and Seppo Laaksonen. 1991. "Weighting via Response Modeling in the Finnish Household Budget Survey." *Journal of Official Statistics* 7:1325–37.
- Goyder, John. 1985. "Face-to-Face Interviews and Mail Questionnaires: The Net Difference in Response Rate." *Public Opinion Quarterly* 49:234–52.
- Groves, Robert, and Mick Couper. 1998. *Nonresponse in Household Interview Surveys*. New York: John Wiley.
- Groves, Robert, Floyd Fowler, Mick Couper, James Lepkowski, Eleanor Singer, and Roger Tourangeau. 2004. *Survey Methodology*. New York: Wiley.
- Groves, Robert, Stanley Presser, and Sarah Dipko. 2004. "The Role of Topic Interest in Survey Participation Decisions." *Public Opinion Quarterly* 68:2–31.
- Groves, Robert, Eleanor Singer, and Amy Corning. 2000. "Leverage-Saliency Theory of Survey Participation: Description and an Illustration." *Public Opinion Quarterly* 64:299–308.
- Groves, Robert, and Douglas Wissoker. 1999. "Early Nonresponse Studies of the 1997 National Survey of America's Families." Assessing the New Federalism Report no. 7. Urban Institute: Washington, DC.
- Groves, Robert, Douglas Wissoker, Liberty Greene, Molly McNeeley, and Darlene Montemarano. 2001. "Common Influences on Noncontact Nonresponse across Household Surveys: Theory and Data." Unpublished manuscript.
- Heberlein, Thomas, and Robert Baumgartner. 1978. "Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature." *American Sociological Review* 43:447–62.
- Hilgard, Ernest, and Stanley Payne. 1944. "Those Not at Home: Riddle for Pollsters." *Public Opinion Quarterly* 8:254–61.
- Izazola-Licea, José, Steven Gortmaker, Victor de Gruttola, Kathryn Tolbert, and Jonathan Mann. 2000. "Assessment of Non-Response Bias in a Probability Household Survey of Male Same-Gender Sexual Behavior." *Salud Publica de Mexico* 42:90–91.
- Kalton, Graham. 1981. *Compensating for Missing Survey Data*. Ann Arbor, MI: Institute for Social Research.
- Keeter, Scott, Carolyn Miller, Andrew Kohut, Robert Groves, and Stanley Presser. 2000. "Consequences of Reducing Nonresponse in a National Telephone Survey." *Public Opinion Quarterly* 64:125–48.
- Kennickell, Arthur, and Douglas McManus. 1993. "Sampling for Household Financial Characteristics Using Frame Information on Past Income." *Proceedings of Survey Research Methods Section of the American Statistical Association*, pp. 88–97. Alexandria, VA: American Statistical Association.
- Lessler, Judith, and William Kalsbeek. 1992. *Nonsampling Error in Surveys*. New York: Wiley.
- Lievesley, Denise. 1988. "Unit Non-Response in Interview Surveys." Unpublished working paper, Social and Community Planning Research, London.
- Lin, I-Fen, and Nora Schaeffer. 1995. "Using Survey Participants to Estimate the Impact of Non-participation." *Public Opinion Quarterly* 59:236–58.
- Little, Roderick. 1982. "Models for Nonresponse in Sample Surveys." *Journal of the American Statistical Association* 77:237–50.
- Little, Roderick, S. Lewitsky, Steven Heeringa, James Lepkowski, and Ronald Kessler. 1997. "Assessment of Weighting Methodology for the National Comorbidity Survey." *American Journal of Epidemiology* 146:439–49.
- Little, Roderick, and Donald Rubin. 2002. *Statistical Analysis with Missing Data*. 2d ed. New York: Wiley.
- Little, Roderick, and Sonya Vartivarian. 2003. "On Weighting the Rates in Non-Response Weights." *Statistics in Medicine* 22:1589–99.
- Lundström, Sixten, and Carl-Erik Särndal. 1999. "Calibration as a Standard Method for Treatment of Nonresponse." *Journal of Official Statistics* 15:305–27.
- Luppes, Martin. 2000. "A Content Analysis of Advance Letters from Expenditure Surveys of Seven Countries." *Journal of Official Statistics* 11:461–80.
- Merkle, Daniel, and Murray Edelman. 2002. "Nonresponse in Exit Polls: A Comprehensive Analysis." In *Survey Nonresponse*, ed. Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, pp. 243–58. New York: Wiley.

- Merkle, Daniel, Murray Edelman, Kathy Dykeman, and Chris Brogan. 1998. "An Experimental Study of Ways to Increase Exit Poll Response Rates and Reduce Survey Error." Paper presented at the annual meeting of the American Association for Public Opinion Research, St. Louis.
- Messonnier, Mark, John Bergstrom, Christopher Cornwell, R. Jeff Teasley, and H. Ken Cordell. 2000. "Survey Response-Related Biases in Contingent Valuation: Concepts, Remedies, and Empirical Application to Valuing Aquatic Plant Management." *American Journal of Agricultural Economics* 83:438–50.
- Nealon, Jack. 1983. "The Effects of Male vs. Female Telephone Interviewers." *Proceedings of the Survey Research Methods Section of the American Statistical Association*, pp. 139–41. Alexandria, VA: American Statistical Association.
- Neyman, Jerzy. 1934. "On the Two Different Aspects of the Representative Methods: The Method of Stratified Sampling and the Method of Purposive Selection." *Journal of the Royal Statistical Society, Series A* 97:558–606.
- Office of Management and Budget (OMB). 2006. *Questions and Answers When Designing Surveys for Information Collections*. Washington, DC: Office of Information and Regulatory Affairs, OMB. Available online at http://www.whitehouse.gov/omb/inforeg/pmc_survey_guidance_2006.pdf (accessed October 25, 2006).
- Pearson, Karl. 1903. "On the Influence of Natural Selection on the Variability and Correlations of Organs." *Philosophical Transactions of the Royal Society, A* 200:1–66.
- Pothoff, Richard, Kenneth Manton, and Max Woodbury. 1993. "Correcting for Nonavailability Bias in Surveys by Weighting Based on Number of Callbacks." *Journal of the American Statistical Association* 88:1197–207.
- Roberts, Paula, Chris Roberts, Bonnie Sibbald, and David Torgerson. 2000. "The Effect of a Direct Payment or a Lottery on Questionnaire Response Rates: A Randomised Controlled Trial." *Journal of Epidemiology and Community Health* 54:71–72.
- Rubin, Donald. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.
- Särndal, Carl-Erik, and Sixten Lundström. 2005. *Estimation in Surveys with Nonresponse*. New York: Wiley.
- Schuman, Howard, and Jean M. Converse. 1971. "The Effects of Black and White Interviewers on Black Responses in 1968." *Public Opinion Quarterly* 35:44–68.
- Singer, Eleanor. 2002. "The Use of Incentives to Reduce Nonresponse in Household Surveys." in *Survey Nonresponse*, eds. Robert M. Groves, Dan A. Dillmon, John L. Eltinge, and Roderick J. A. Little, pp. 163–77.
- Singleton, Royce, and Bruce Straits. 2005. *Approaches to Social Research*. 4th ed. New York: Oxford University Press.
- Smith, Fred. 1983. "On the Validity of Inferences from Non-Random Samples." *Journal of the Royal Statistical Society, Series A* 146:394–403.
- . 1994. "Sample Surveys, 1975–1990: An Age of Reconciliation?" *International Statistical Review* 62:3–34.
- Stephan, Frederick, and Philip McCarthy. 1958. *Sampling Opinions*. New York: Wiley.
- Stephenson, C. Bruce. 1979. "Probability Sampling with Quotas: An Experiment." *Public Opinion Quarterly* 43:477–96.
- Valliant, Richard. 1993. "Poststratification and Conditional Variance Estimation." *Journal of the American Statistical Association* 88:89–96.