Federal Committee on Statistical Methodology

Best Practices for Nonresponse Bias Reporting

Prepared by the Nonresponse Bias Subcommittee of the

Federal Committee on Statistical Methodology

June 2023

Recommended citation: Madans JH, Earp MS, Blumberg SJ, Christopher EM, Thompson KJ, Sivinski RG, Frenk SM, Fakhouri THI, Piscopo KD. 2023. *Best Practices for Nonresponse Bias Reporting*, FCSM-23-01 Nonresponse Bias Subcommittee, Federal Committee on Statistical Methodology. June 2023.

Federal Committee on Statistical Methodology

Members

Ann Carson
Bureau of Justice Statistics

Chris Chapman
National Center for Education Statistics

Jennifer Childs
U.S. Census Bureau

Tala Fakhouri
U.S. Food and Drug Administration

John Finamore
National Center for Science and Engineering Statistics

Ellen Galantucci
Federal Maritime Commission

Michael Hawes
U.S. Census Bureau

Travis Hoppe
National Center for Health Statistics

Wendy Martinez
Bureau of Labor Statistics

Lisa Mirel
National Center for Health Statistics

Jennifer Nielsen
National Center for Education Statistics

Jennifer M. Ortman
U.S. Census Bureau

Anne Parker
Internal Revenue Service

Jennifer Parker
National Center for Health Statistics

Mark Prell (Co-Chair)
Economic Research Service

Michael Ratcliffe
U.S. Census Bureau

Rolf Schmitt
Bureau of Transportation Statistics

Darius Singpurwalla
National Center for Science and Engineering Statistics

Bob Sivinski (Chair)
Office of Management and Budget

Matt Spence
U.S. Census Bureau

Ed Strocko
Bureau of Transportation Statistics

G. David Williamson
Centers for Disease Control and Prevention

Linda Young
National Agriculture Statistic Services

Lori Zehr (Secretary)
U.S. Census Bureau

Nonresponse Bias Subcommittee

Members

Stephen Blumberg (Co-Chair)
National Center for Health Statistics

Elise Christopher
National Center for Education Statistics

Morgan Earp (Co-Chair)
National Center for Health Statistics

Tala Fakhouri
U.S. Food and Drug Administration

Steven Frenk
National Institutes of Health

Jennifer Madans (Co-Chair)
National Center for Health Statistics

Kathy Downey Piscopo
Substance Abuse and Mental Health Services Administration

Bob Sivinski
Office of Management and Budget

Jenny Thompson
U.S. Census Bureau

Table of Contents

OBJECTIVES

The goal of this report is to provide an overview of best practices for discussing and reporting on nonresponse bias in estimates obtained from survey data. Nonresponse bias is one of the main threats to data quality in federal surveys, as discussed in the 2020 report *A Framework for Data Quality* from the Federal Committee on Statistical Methodology (FCSM). Yet there are no consistent standards for reporting on nonresponse bias, a problem clearly demonstrated in another 2020 FCSM report: *A Systematic Review of Nonresponse Bias Studies in Federally Sponsored Surveys* (Miller et al., 2020). The current report addresses the latter concern.

This report and both of the 2020 FCSM reports elaborate on the issues included in the 2006 Office of Management and Budget (OMB) guidance *Standards and Guidelines for Statistical Surveys*. The OMB guidance focuses on several aspects of data quality and provides standards for describing surveys and reporting on nonresponse. The OMB guidance requires that all federal surveys conduct a nonresponse bias analysis if the program expects a unit-level response rate less than 80 percent, or an item-level response rate less than 70 percent (Office of Management and Budget 2006). The current report offers additional guidance for reporting on the methods, results, and conclusions from those nonresponse bias analyses.

The literature on how to conduct nonresponse bias analyses is extensive and growing. The methods used for any data collection will be a function of the characteristics of that data collection and the availability of information to evaluate bias. The intent of the current report is not to provide a comprehensive guide to conducting nonresponse bias analyses. Rather, it is to provide guidance on reporting on the analyses done. To do this, the current report sometimes refers to various methods for conducting, and aspects of, nonresponse bias analyses. But these methods are not fully described. Readers will need to refer to the many existing references for more information on these methods.

The current report also does not offer guidance for assessing whether the nonresponse bias present in any data collection is small enough to be acceptable or so large as to be unacceptable. Rather, the guidance in the current report is intended to yield sufficient documentation of the methods and results of nonresponse bias analyses so that readers can

determine for themselves whether the data are of sufficient quality — that is, the data "are fit for their intended uses in operations, decision making, and planning" (Redman, 2001, p. 74). Survey programs generally cannot predict all purposes for which their data will be used. Moreover, even within the context of a specific intended use, assessment of "fitness for use" requires a multidimensional approach that considers many elements of data quality (FCSM, 2020). Nonresponse bias is just one (albeit important) element in one of the FCSM data quality dimensions (accuracy and reliability).

The guidance presented in the current report is designed for survey data collections. However, many of the best practices in this report can also be applied to reporting analyses of bias for other data sources. For example, gaps in data resulting from malfunctioning sensors in a large array can be thought of as equivalent to nonresponse in sensor-based data collections. Item-level missingness in administrative data may be analogous to item-level missingness in surveys, especially when items are not mandatory for completion of the form.

OVERVIEW

The generalizability of survey estimates, a key quality indicator (FCSM, 2020), relies on the premise that information is obtained from the entire universe or a randomly selected, representative sample of potential survey respondents selected from the population of interest. If information is not complete, response could be disproportionate across groups. Nonresponse bias can occur when the likelihood of responding to surveys or other data collection methods (or items) is correlated with key data collection outcomes. Put another way, nonresponse bias can occur when respondents differ from nonrespondents on key characteristics and key outcomes. Nonresponse bias affects both the accuracy and the reliability of population estimates.

Nonresponse can occur for a variety of reasons both at the sample unit level (e.g., (persons, establishments, schools, farms, or hospitals) and at the item level (e.g., variables). Respondent knowledge and perceptions of burden, saliency, sensitivity, and/or privacy threats may vary across sampled units, leading to nonresponse that is non-random. When those and other drivers of nonresponse are related to the key survey items or characteristics being measured by the survey, nonresponse bias can occur.

The relationship between the unit-level response rate and nonresponse bias is complex and researchers should avoid relying on response rates alone to determine data quality. Several studies have shown that response rates are not always correlated with nonresponse bias (Groves & Peytcheva, 2008; Hendra & Hill, 2019) and that boosting the response rate by bringing in additional cases does not necessarily decrease bias (Heerwegh, Abts, & Loosveldt, 2007). Some studies have even shown that efforts to increase response rates can increase bias by recruiting subgroups that were already at a higher propensity to respond, versus those at lower propensities to respond, which could further exacerbate the nonresponse bias (Merkle & Edelman, 2009) or

increase measurement error by coercing unwilling participants that give low quality responses (Barge & Gehlbach, 2012; Groves, 2006). It should also be noted that while response rates are often used as a proxy of overall data quality, using the response rate alone ignores all other possible sources of error that can cause survey estimates to differ from the underlying population parameters, such as coverage error, sampling error, measurement error, and processing error.

Survey programs invest a significant amount of time and resources trying to maximize response rates and minimize the risk of nonresponse bias. While response rates alone are not indicative of data quality, the lower the response rate, the greater the opportunity is for nonresponse bias to occur (Groves, 2006; Groves & Peytcheva, 2008). This can be especially true in establishment surveys, where the impact of big retailers or producers not responding can have a large impact on the final estimate.

An overarching best practice is to plan on conducting a nonresponse bias analysis as part of the survey life cycle. Besides setting aside time for this research and analysis, it is important to include questions in the data collection instrument that might help in the evaluation of nonresponse bias. For example, if a survey includes a longitudinal component, it may be useful to include questions in an early round that can be used to evaluate nonresponse during later follow-up data collection. Advice from program staff and contractors with subject matter and survey methods expertise should be solicited when developing the plan for the nonresponse bias analysis. Resources need to be devoted to finding comparable and high-quality auxiliary data to aid in the process. A useful nonresponse bias analysis is motivated by a desire to explore relationships between survey nonresponse – both at the unit level and at the item level – and the quality of the estimates produced.

Nonresponse bias analysis should be viewed as part of the core objective to mitigate total survey error — the difference between a survey estimate and the true value— thus enhancing the quality of the data. This difference results from the accumulation of all errors that arise in the design, collection, processing, and analysis of survey data (Biemer, 2010). Mitigation of response bias can be done during data collection using design elements (such as responsive design) or experiments that target known or hypothesized sources of error. Mitigation can also be done as part of the post-data collection processing where efforts are made to reduce, to the extent possible, any bias identified in the nonresponse bias analysis, for example by adjusting survey weights.

The nonresponse bias analysis report should provide a clear description of the various analytic approaches used to determine the level of bias and any mitigation strategies implemented. As post-data collection analysis and mitigation generally are done in stages, the report should document each stage of analysis and mitigation. An ascertainment of residual bias after all mitigation strategies, if any, have been implemented is required for a final determination of potential bias and how it might affect the use of the data. This information enables data users to determine whether a particular data source is fit for their analytic purpose and for the users of statistical and analytic products to evaluate the appropriateness of findings.

While this report focuses on the discussion and reporting of nonresponse bias in surveys, nonresponse bias can even occur in the external sources used to evaluate potential nonresponse bias in surveys. Given that the sources used to evaluate nonresponse may also suffer from nonresponse bias, nonresponse rates and nonresponse bias should be reported for both the survey being evaluated for nonresponse bias and (if possible) the data used to evaluate nonresponse bias, to assess the potential for compounded or distorted bias estimates.

Just as planning a nonresponse bias analysis needs to be built into the survey life cycle, successful mitigation strategies should be implemented for future data collections; therefore, recommendations for mitigating nonresponse bias based on the results of the study should be included in the report and should accompany the release of the data. The knowledge gained from conducting a nonresponse analysis should ultimately be used to mitigate future nonresponse, whether it is through design elements or post-survey adjustment methods (Wagner et al., 2012; Lepkowski et al., 2013). For recurring surveys, the findings of the nonresponse bias analysis can be used to target future data collection resources more efficiently and/or improve the existing nonresponse bias mitigation strategies (e.g., weighting or imputation).

While survey programs often devote resources to assessing and addressing nonresponse bias, the results of this work are often not shared or published for public consumption. This creates two problems: 1) it prevents data users from understanding data quality concerns and biases; and 2) it prevents other survey programs from learning from each other's experiences, including both successes and challenges.

The best practices – and associated guidelines – presented in this document address two major concerns highlighted by the FCSM in 2020: 1) the threat of nonresponse to data quality, specifically accuracy and reliability (FCSM, 2020); and 2) the lack of standardized reporting and evaluation of nonresponse bias (Miller et al., 2020). This report provides a framework for addressing these problems. Reporting and disseminating nonresponse bias analysis and research conducted by federal agencies and survey researchers more widely within a standard framework allows data collection entities to better plan, share, and evaluate nonresponse bias research and results within and across the user community.

The report is structured around seven best practices that provide guidance for developing a nonresponse bias analysis report: 1) describing the survey, 2) providing unit-level response rates and discussing the potential for nonresponse bias, 3) describing the plan for evaluating, quantifying, and mitigating nonresponse bias, 4) describing and justifying auxiliary data used to assess nonresponse bias, 5) describing the results of the nonresponse bias analysis, 6) summarizing the major conclusions of the nonresponse bias analysis, and 7) discussing recommendations for data collection methods and adjustment strategies to mitigate nonresponse bias in future data collections. The best practices are drawn from previous research and build on recommendations introduced in the FCSM report *Transparent Reporting for Integrated Data Quality: Practices of Seven Federal Statistical Agencies* (Prell et al., 2019). The best practices reinforce and are consistent with the current focus on the need to appropriately document all data collection procedures, including comprehensive metadata. As is the case with all guidance frameworks, not all best practices will apply in all situations. However, all best practices should be considered when developing and reporting on nonresponse bias and there should be appropriate justification for not following any best practice.

GUIDANCE FOR DEVELOPING A NONRESPONSE BIAS REPORT

Best Practice 1: Describe the survey that is the subject of the nonresponse bias analysis

There are several key elements of a survey that impact data quality (e.g., accuracy and reliability) and the magnitude of total survey error (e.g., coverage error, measurement error, nonresponse error, etc.) including the overall construct, design, administration, target population, sampling frame, sampling methods, and data collection modes. These elements can all impact data quality and should be discussed and considered before attempting to assess nonresponse bias. According to the FCSM report, *A Framework for Data Quality* (2020), "The accuracy of a data product reflects the accuracy of the input sources and all processing and calculations performed to transform those data into outputs. As these errors can accumulate throughout the data lifecycle, the accuracy of a data product will reflect a combination of the accuracy of its input sources, the processing steps applied to those inputs, and any additional calculations performed to transform the data into outputs" (p. 17). Given the potential for errors even without nonresponse, it is important to first evaluate the validity and generalizability of the key survey estimates given the overall survey design. The report should describe the key aspects of the survey design up through data collection, such as the population of inference, target population, sample frame, potential for coverage error, sample design, sample unit, data collection modes, key survey items, and any nonresponse bias mitigation used during data collection.

Guideline 1.1. Describe the survey target population

The report should describe the specific target population (people or establishments) of the survey and the population for which inferences will be made. For example, the target population for a survey might be the U.S. non-institutionalized population aged 12 or older, while other surveys may target agricultural operations, business establishments, or educational institutions.

Guideline 1.2. Describe the survey sample frame

Probability samples utilize a frame (i.e., a finite list of sampling units in the population), where each sample unit has a known probability of selection. For example, an address-based household survey might construct a frame using the Census Bureau's Master Address File, or by sending staff into the field to make lists of addresses (area listing), or by purchasing a commercial list of household addresses. Telephone-based household surveys may draw from banks of consecutive telephone numbers known to include at least one directory-listed residential telephone number or cell phone number. In some cases, a list of email addresses may provide a useful frame when all units have an email address and have access to email (e.g., all enrolled students at a university). For example, the Bureau of Labor Statistics uses unemployment insurance records as the sample frame when surveying businesses on job openings and labor turnover.

The report should describe the survey frame and how it was built. Some surveys might rely on multiple frames in which case information on each frame used and how they are used together should be discussed. The report should provide a definition of the sampling units and clearly define how units are classified as eligible versus ineligible or out of scope.

Guideline 1.3. Discuss the potential for coverage error of the frame

Coverage error represents the gap between the true value given the target population and the estimate given the sampling frame. According to the FCSM report, *A Framework for Data Quality* (2020), "For sample surveys, coverage error occurs when the sampling frame differs from the target population. Substantial coverage errors affect the utility of the data for inferences about the target population" (p. 38). For example, target household units without landline

telephones would not be included in a frame consisting of landline phone numbers and would therefore have zero probability of being sampled in a survey of households.

Available sampling frames do not always align with the target population and may exclude certain populations and/or may not be complete or accurate. Even when the scope of a sampling frame aligns perfectly with the target population, it may include sample units that are no longer eligible or be missing newer sample units (e.g., recently formed households, schools, or business) that are eligible. Frames used for population or establishment surveys may also be out of date by the time the sample is selected if new housing units or establishments are formed between the time the frame is created and data collection begins. Any issues that impact the generalizability of the sample to the target population should be described for transparency and credibility.

Guideline 1.4. Describe the sample design

The report should provide a description for how the sample units were selected from the sample frame. Some surveys use simple random samples, while larger surveys meant to generalize to the U.S. population and have more disaggregated geographic levels (such as the state level) might use multi-stage stratified or cluster designs. If stratification or clustering was used, the report should provide a discussion of how these sampling decisions impact the estimates. Descriptions of any unequal probabilities of inclusion used should be included.

Guideline 1.5. Describe the sample unit

The report should describe the level(s) at which data were collected. For example, the report should discuss whether the survey sampled entire households as the unit or sampled persons within a sampled household, or both; whether the survey sampled entire educational institutions as the unit or sampled classes within sampled institutions and then students within

sampled classes, or at all levels; or whether the survey sampled the entire business as the unit or sampled one or more locations for a given business.

Guideline 1.6. Describe the survey data collection modes

The report should describe the methods used to collect the data, particularly whether telephone, web, computer-assisted interviewing (CAI), touch tone or interactive voice response (IVR), data exchange, or a combination of modes (mixed mode collection) were used. Data collection mode may be correlated with coverage error, measurement error, and response error. For example, households without telephones cannot participate in telephone-only surveys (coverage error), whereas face-to-face data collection could potentially result in higher response rates (and perhaps lower nonresponse error) but also introduce measurement error if respondents are less likely to respond honestly to sensitive questions that are interviewer administered in person.

Guideline 1.7. Describe the survey's key survey items, and identify those that will be used in the nonresponse bias analysis

While a survey may collect a variety of data elements, each survey is designed to produce specific key estimates or outcomes. The report should provide a description of the survey's key estimates or outcomes and the set of variables used to produce them. Any other survey variables that are assessed in the nonresponse bias analysis should also be described.

Guideline 1.8. Describe any nonresponse mitigation strategies tested during data collection

The report should document and describe specific and targeted activities implemented or tested during data collection to increase response rates and reduce potential bias. Examples of such activities include experiments with incentives, respondent materials, or contact protocols.

While all surveys use a variety of methods to increase response rates, this guideline calls specifically for specifying new, revised, or experimental data collection protocols designed to reduce nonresponse bias and for which an evaluation of bias reduction is warranted.

Best Practice 2: Provide unit-level response rates for the survey and discuss potential for nonresponse bias

Unit-level response rates are unit-level performance metrics that measure the proportion of the eligible sample units that responded to a survey. The report should provide a complete definition of what constitutes a "response." While obtaining responses to all appropriate questions is the goal, breakoffs where the respondent completed a predetermined subset of questions that address the key elements of the survey are often considered as having responded to the survey. The criteria used to determine when a partial response is considered a response to the survey should be described.

The unit-level response rate is often used as a proxy measure of survey data quality; however, it is not necessarily a strong predictor of the nonresponse bias in survey estimates (Groves & Peytcheva, 2008; Hendra & Hill, 2019). While a high unit-level response rate is generally consistent with low bias, and a lower unit-level response rate is more likely to be associated with bias, there are some situations where the latter may not lead to bias. For example, response rates are associated with bias only if the likelihood of response is related to the key survey items. It is possible for a survey to have a low response rate but also key measures that are not related to survey response propensity. Therefore, those measures will not be subject to nonresponse bias. The unit-level response rate is the starting point for evaluating nonresponse bias. Conducting nonresponse bias analysis and reporting on the results is particularly important if unit-level rates are low (e.g., below 80 percent; Office of Management and Budget, 2006).

Guideline 2.1. Identify and report unit-level response rates based on weights that adjust for selection probabilities

The unit-level response rate measures the proportion of eligible units that responded to the survey. Unit-level response rates computed without survey weights provide a useful metric for evaluating the success of data collection for the selected units. But as a metric for evaluating the success of the survey with respect to the population sampled, the response rate calculations should incorporate the base weight for each unit. The base weight often is the inverse probability of selection and its inclusion in the response rate estimate results in a response rate calculation that uses a denominator that is the approximate total number of eligible sample units in the frame (population total). For establishment surveys, a measure of size should be used to calculate the weighted response rate to reduce the influence of small units with large sampling weights and appropriately represent the anticipated contribution of the large establishments on survey estimates.

Unit-level response rates should be reported, to the extent possible, using standard definitions such as those published by the American Association for Public Opinion Research (AAPOR, 2023). Standard definitions have been vetted by the statistical and survey research communities and the justifications for the definitions are transparent and available to the user community. Rates that are calculated using standard definitions can also be compared over time for the same data collection as well as across data collections.

Guideline 2.2. Report unit-level response rates for geographic areas or key subgroups for which population-based estimates are published

Examining the unit-level response rates of subgroups can help identify potential sources of nonresponse bias, especially if it is found that the survey estimates vary by subgroup. The

subgroups examined should reflect the subgroups for which key estimates are reported. For example, if the survey provides estimates for Census regions, lower levels of geography, or key subgroups, then response rates should also be provided for these subgroups. In addition, it can be instructive to produce response rates for subgroups that historically have lower response rates. Variation in response rates for subgroups not only identifies potential sources of bias but also alerts the user to potential limitations of the data set for that user's intended analytic objectives. This variation can also be an important consideration when evaluating data through an equity lens.

The ability to calculate response rates for subgroups will be limited by the data available at the sample level. Without a rich sampling frame or matched auxiliary data, characteristics of the nonrespondents may be unknown. Subgroup-specific response rates should be reported for key subgroups where information is available to directly calculate the response rate.

Coverage ratios — a comparison of the sum of the sampling weights from the sampled units that have a particular characteristic to the same estimate (e.g., number of people with that characteristic) from an independent source — are sometimes used as a proxy for reporting response rates by key subgroups, when the information needed to identify subgroups is not available at the sample level. It should be noted, however, that coverage ratios calculated using the base weights conflate nonresponse error with coverage error. If coverage error has been described well (Guideline 1.3), it may be possible to use coverage ratios to infer differences in response rates by key subgroups.

R-indicators and partial R-indicators discussed in Schouten et al (2012) and the balance and distance indicators discussed in Lundquist and Särndal (2013) are additional methods of evaluating variation in unit-level response rates for subgroups.

Guideline 2.3. Report item-level response rates for all key survey items

The report should discuss the item-level response rates for all key survey items (see Guideline 1.7). A survey might have a very high unit-level response rate, but particular items might have high item-level nonresponse rates, therefore rendering the item either unusable or needing imputation. Reporting item-level response rates is particularly important if some partial responses have been classified as unit-level respondents. For example, some surveys classify units as respondents if they just respond to one or more key survey items, which can lead to high item-nonresponse after the cutoff for defining when a partial response is a response to the survey.

The item-level response rate calculations should incorporate the base weight for each unit. Unweighted and weighted item-level response rates may both be considered, but weighted rates are better indicators of potential bias. The unweighted rate describes the extent to which sample units provided information and reflects the success or challenges of the field work. High unweighted item-level nonresponse might indicate that the question was too sensitive, or the cognitive burden required to respond was too high, or the question was unclear, or respondents simply don't have the answer. This may have implications for planned analyses or future surveys. However, the weighted rate is assumed to provide a better indicator of how well the reporting units describe the target population as it accounts for unequal probabilities of selection.

Guideline 2.4. Report response rates for each sampling stage that provides actionable information on survey participation and/or coverage

For surveys that employ complex sample designs with more than one stage of selection or clustering, response rates should be reported for each sampling stage and include a discussion of how nonresponse at each stage could affect the bias in the final estimates. For example, school-based surveys sample schools and then students within schools. Address-based surveys sample

geographic areas then sample households within those areas, and persons within households. Establishment surveys sometimes breakdown sampling and unit response into different stages. For example, establishments may first be selected from a frame; then contacted to get mailing or email addresses and to verify frame information; then contacted to enroll them in the survey; and then recontacted at the start of data collection. Response rates can vary by phase of data collection, which could have a cumulative or counteracting impact in bias, so it is important to report response rates at each reporting phase, not just during the final data collection phase (see Earp et al., 2018).

Guideline 2.5. Report response rates for each wave of data collection in longitudinal surveys

Attrition at each wave of a panel or longitudinal survey can affect nonresponse bias as data collection continues. Response rates should be included for each data collection wave starting with the weighted and unweighted response rates for the first wave of data collection. Response rates should be reported for all key items in all waves (see Guideline 1.7). Response rates across waves can increase or decrease over time, and the variation in response rates can be related to key items. It should be noted whether cases can be missing for a given wave but still be considered part of the panel, or if they are removed following nonresponse at a certain wave, as this affects response rate calculations for subsequent waves.

Guideline 2.6. Report response rates for each group when employing responsive design or treatment groups

Adaptive or responsive designs intended to increase response from "hard to reach" units (i.e., population groups) are often used to reduce nonresponse bias (Wagner et al., 2012; Lepkowski et al., 2013). For example, treatments may be applied to all sampled units or subgroups to increase response (e.g., monetary incentives may be offered for completion of a

survey), but not all sampled units may be offered the same treatment. However, these efforts can have unintended consequences and inadvertently introduce or increase bias. Response rates by treatment group (i.e., monetary offer amount) should be reported.

Guideline 2.7. Discuss the potential for nonresponse bias given weighted response rates, response rates by sampling stage, waves of data collection, and key subgroups

The potential bias introduced by overall nonresponse and by nonresponse for each of the key items should be discussed. Differential nonresponse by subgroup can affect bias in overall findings as well as findings for that subgroup, so the potential bias for subpopulations of interest should also be discussed. Level and observed variation in response rates by sampling stage and waves of data collection should also be discussed to present a more complete picture of how the variation may be affecting estimates. Weighted response rates should be reported and discussed in terms of how bias could be affected. (See Guidelines 2.1 through 2.6).

This discussion is intended to offer predictions and hypotheses to be evaluated and quantified when the analytic plan (Best Practice 3) is implemented.

Best Practice 3: Describe the plan for evaluating and quantifying nonresponse bias and any mitigation strategies to be employed

The report should include a clear description of activities used to reduce nonresponse bias during data collection (see Guideline 1.8), the analytic plan for evaluating nonresponse bias in the final data file, and, if applicable, the post-data-collection mitigation strategies (see Guideline 3.3). The description of the plan should include the order in which analysis and mitigation are done. Mitigation strategies are those used to reduce biases identified in the nonresponse bias analysis.

There are a variety of evaluation methods that can be used to assess nonresponse bias. The nonresponse bias assessment method(s) selected should depend on data collection characteristics and the availability of appropriate auxiliary sources of information (see Best Practice 4). To the extent possible, different methods should be used as each method focuses on different aspects of the effects of nonresponse and the use of multiple methods provides a more complete evaluation of bias. There is no standard or preferred order or manner in which different methods or strategies should be combined; however, the analysis and mitigation plan should provide the order of evaluation activities used as the effects of the methods or strategies used are cumulative. The report should help the reader understand the process through with mitigations were applied. New approaches to nonresponse bias analyses are being developed, so it is critical to keep informed on advances in the field.

Groves (2006) and Wagner (2012) both suggest five main approaches to assessing nonresponse bias:

1) assessing response rates over waves/time, across surveys, and within subgroups;

2) comparing estimates across surveys or to similar estimates from other sources;

3) studying estimate variation within the existing survey by comparing estimates within surveys across waves/time or via nonresponse follow-up studies;

4) comparing statistics within a survey using rich sampling frame or matched auxiliary data; and

5) assessing the impact of mitigation strategies by monitoring data collection and contrasting alternative postsurvey adjustment methods.

These standard nonresponse bias analysis approaches fall into two main categories – those that use information from the data collection itself and those that use auxiliary data.

The report should include a complete discussion of all evaluation methods used, including the rationale for the selection of each method along with an explanation of why some standard methods are not used. In addition, the discussion should address the aspect of nonresponse that the method will address.

Guideline 3.1. Describe and justify all nonresponse bias evaluation methods used

Evaluation methods can be direct or indirect. Direct assessment methods use data from the data collection such as the sampling frame or directly matched auxiliary data. Indirect assessment methods compare estimates from the data collection to similar estimates from other sources. The report should identify and describe all methods used to evaluate nonresponse bias, identifying which are direct and which are indirect.

Limitations of different methods are discussed by both Groves (2006) and Wagner (2012). The report should compare and contrast the limitations associated with all selected nonresponse bias assessment methods. The report should also discuss any steps taken to address these limitations and how the varying evaluation methods may or may not complement each other.

Guideline 3.2. Describe the analysis plan for comparing and assessing nonresponse bias levels across key stages and subgroups

The report should discuss all aspects of the plan for comparing and assessing nonresponse. This should include the potential impact on nonresponse bias of any observed differential response rates for multiple stages of data collection if applicable or for subgroups of

interest (see Guideline 2.2). The plan for determining how the level and variation in response rates could impact bias should also be discussed.

Guideline 3.3. Describe post data collection nonresponse bias mitigation strategies that were employed and how they might reduce nonresponse bias

The report should describe all the strategies used after data collection was completed to mitigate previously identified biases, including adjustments to sample weights and imputation. Sufficient detail on any new post-data collection mitigation methods developed should be provided so that readers will understand the methods used.

Base weights can be adjusted in a variety of ways to mitigate nonresponse bias. Descriptions of unit-level nonresponse bias mitigation strategies such as poststratification, raking, and adjustment cell weighting should include an explanation of the models used, underlying assumptions about missingness, and discussion of the alternative approaches evaluated. These adjustments often involve a tradeoff between variance and bias. To the extent possible, decision criteria for managing these tradeoffs should be established prior to decisions about whether to employ mitigation strategies. Any efforts to attenuate the increased variance resulting from bias mitigation, such as weight trimming, should be clearly described.

Best Practice 4: Describe and justify all available sources of auxiliary data used in the nonresponse bias analysis

Auxiliary data are broadly defined as data that contain variables on the target population that are the same as or at least correlated to the key survey items. The type of nonresponse bias analysis as well as the type of inferences that can be drawn depend on what auxiliary sources are available at the time of analysis. Approaches that employ auxiliary data can provide critical information on nonresponse bias and identify ways to mitigate it. Because of the potential

importance of auxiliary data for the evaluation and mitigation of nonresponse bias, issues related to the use of auxiliary data should be addressed. For example, the availability, completeness, timeliness, and accuracy of auxiliary data will impact the methods selected, how the analysis is done, and the interpretation of the results.

The auxiliary data can be obtained from external sources or may include data collected through the survey but not used for reporting purposes (e.g., paradata). Auxiliary data fall into three broad categories:

1. Unit level: variables from another source that match (one-to-one) to the survey's sample or collection units for respondents and nonrespondents. Unit-level data can come from a variety of sources including the survey frame, another survey, administrative records, paradata, or commercial data.
2. Area level: reliable aggregate totals from subgroups or domains from another source that can be linked (many-to-one) to the unit-level survey data.
3. Benchmarks: reliable aggregate totals from another source that are closely related to the survey's respondent-based estimates.

As previously noted, auxiliary data that can be used to evaluate and mitigate bias must be at least weakly correlated to key survey outcomes. Auxiliary data that include the same concepts for the same target population provide the most direct assessment of nonresponse bias. Auxiliary data with high correlations can be used as "proxies" for key outcomes. But even variables that are weakly correlated may also be useful for assessing nonresponse bias. Groves (2011), Krueger and West (2014), Johnson and Smith (2017), Lohr (2004), and Rao (2003) suggest using multiple sources of auxiliary data whenever possible. The strength of the relationships between the auxiliary data and survey characteristics should be reported (Guideline 4.3).

See Appendix A for more information on selected types of auxiliary data. Advances in the availability and diversity of big data and in the development of artificial intelligence will likely lead to the identification and increasing use of new types of auxiliary data, including data extracted from maps, images, web scraping, and more. Any effort to develop a complete list of all possible types of auxiliary data would be quickly outdated.

Guideline 4.1. Describe all sources of auxiliary data used for the nonresponse bias analysis

Depending on the survey being assessed for nonresponse bias, there may be multiple sources of auxiliary data or very few, if any. All major potential auxiliary data sources should be discussed. The discussion of the auxiliary data should include whether they are available for all sampled units or for the entire population (Brick, 2013; Lohr, 2004; Rao, 2003; Wagner, 2012). If unit-level data are available for all sampled units, then quantified nonresponse bias at the unit level can and should be reported (Brick, 2013; Groves, 2006; Krueger & West, 2014; Rao, 2003; Wagner, 2012). Even if data are not available for all sampled units, geographic and demographic information should be used when available, and should be used to report bias at the area level or subdomain level (Brick, 2013; Groves, 2006, Rao, 2003; Wagner, 2012). In addition, if current population totals (benchmarks) are known, these should be compared to survey estimates and included in the report.

Sources of auxiliary data can be obtained from the survey itself. Paradata (e.g., how long the interview took to complete or how long the respondent took to fill out the form) may be actively or passively collected during the data collection process. Standardized interviewer observations might be tailored such that they are expected to be related to the key survey variables (West, 2013; West & Kreuter, 2015; West & Little, 2013). In addition, in surveys that screen for eligibility, adding questions to the screening instrument may provide auxiliary data

that permit evaluation of nonresponse bias at the post-screening stage of recruitment (Wagner et al., 2012; Montaquila et al., 2013).

The report should describe the sources of the auxiliary data selected. In addition, information on potential sources evaluated but not selected should be provided along with criteria for selecting the sources used. All appropriate auxiliary sources should be used in the evaluation. If multiple sources of auxiliary data exist, the report should discuss nonresponse bias results using each data source. It is expected that sources will vary in terms of quality and relevance, and that results from different auxiliary sources may vary. See Appendix A for a discussion of different auxiliary data sources.

Guideline 4.2. Discuss the characteristics, quality, and completeness of auxiliary data in terms of coverage, item missingness, measurement error, and timeliness

The importance of reliable and conceptually comparable auxiliary data for all key survey items in nonresponse bias analysis is emphasized in the survey research methods literature; however, this rarely occurs in practice (Brick, 2013; Andridge & Thompson, 2015; Olson, 2013; Groves, 2006). A good source to review for assessing and reporting auxiliary data quality is the FCSM *Data Quality Assessment Tool for Administrative Data* (Iwig, Berning, Marck, & Prell, 2013).

There are a variety of challenges and limitations to consider when using auxiliary data. The availability of auxiliary data may be very limited. The auxiliary data may also include nonresponse bias and measurement error. Ideally, auxiliary variables should provide non-missing values for all respondents and nonrespondents and be free of measurement error (Kreuter at al., 2010; Olson, 2013; Massey & Tourangeau, 2013). This ideal is difficult to achieve. Therefore, the nonresponse bias analysis report should discuss the strengths and limitations of each

auxiliary data source with the respect to the analysis. The most useful auxiliary data are themselves of high quality with low levels of missingness and measurement error. See Appendix B for a discussion of missing data, coverage, measurement error, and data timeliness.

Guideline 4.3. Discuss the relationship between auxiliary data being used to evaluate nonresponse bias and the key survey items

When assembling different sets of auxiliary data for nonresponse bias analyses, priority should be given to identifying variables of high quality (see Appendix B) that are expected to be related both to survey response propensity and the survey characteristics of interest (i.e., key survey items) (Andridge & Thompson, 2015; Groves, 2006; Krueger & West, 2014; Olson, 2013; Särndal & Lundström, 2008; Kreuter et al., 2010; Massey & Tourangeau, 2013). Of the two, it is more important to find strong correlates of survey items in nonresponse analysis. Several studies show that using auxiliary variables that are weakly correlated with the survey variables of interest may not only result in poor nonresponse bias assessment and nonresponse adjustment but may also unnecessarily increase variance (if included in the adjustment procedure) (Kreuter et al., 2010; Little & Vartivarian, 2005; Olson, 2013; McConville & Toth, 2019).

While a strong association between related measures is expected, the association may vary depending on differences in definition, collection units, and timeliness. Differences between the survey of interest and the benchmark do not necessarily indicate nonresponse bias but can also be due to other factors. In addition to subtle differences in definitions, reference populations, and timing, survey estimates may also differ from benchmarks because of the survey's "social setting" (i.e., the location of survey administration such as the respondent's home, a clinic, or a school, etc.), mode of collection (e.g., online or paper forms), context or priming effects (from

other questions on the survey or the topic of the survey), and respondent fatigue (more common for questions near the end of the survey), all of which have been shown to influence survey responses in some instances. In many cases, the associations may be much weaker for survey measures that measure similar but related concepts. In addition, all surveys are subject to nonresponse and potential nonresponse bias, as well as measurement, coverage, processing, and sampling errors.

When items in the survey data can be matched to auxiliary data items (e.g., the same variable exists for the same unit in another data source), then it is a best practice to assess the relationship between the items. For example, agricultural production surveys often match variables to similar variables collected on the Census of Agriculture to assess the potential for nonresponse bias. By comparing reported Census values for survey respondents and nonrespondents, it may be possible to determine if there is potential for nonresponse bias. Correlation and regression type analyses are commonly used to assess the relationship between matched variables. Such analyses can indicate the degree to which the auxiliary data might be useful in understanding the amount of bias due to nonresponse in the survey data.

While it is common practice in household surveys to adjust on age, sex, race/ethnicity, and geographic location, relying on these characteristics alone may not be enough to adequately adjust for nonresponse bias. If no other auxiliary data exists, one should at the very least report interaction effects between key demographics in relation to survey nonresponse (Brick, 2013; Peytcheva & Groves, 2009; Phipps and Toth, 2012). For example, using regression trees to model the relationship between establishment or household characteristics and nonresponse can provide insight as to whether nonresponse propensities vary across demographic subgroups. While variation across demographic subgroups alone does not necessarily indicate bias, it can be

informative if inequities in nonresponse are also thought to be related to inequities in the survey outcome.

Best Practice 5: Describe results of nonresponse bias analysis including for all key survey items

The results — that is, the extent of bias identified from the nonresponse bias analysis — should be reported for each method used, even if the results were inconclusive. If pre or post mitigation strategies were implemented, the results of those strategies on bias indicators and key survey items should also be reported. If nonresponse bias analyses reveal that the relationship between response propensity and key survey item values is not well understood and there are concerns about the potential biasing effects of nonresponse, then the report may include a recommendation for future nonresponse follow-up studies. However, the recommendation should include a caveat that these studies should only be undertaken if enough resources can be dedicated to ensuring a study with sufficient sample size and without similarly high rates of nonresponse.

Guideline 5.1. Provide and discuss the results of the nonresponse bias analyses specified in the analytic plan

The report should discuss results of all components of the analysis included in the analytic plan. The discussion should address potential bias using indicators of bias for the key survey items (see Guideline 1.7), for key subgroups (see Guideline 2.2), and for each sampling stage and wave of data collection (see Guidelines 2.4 and 2.5). The results of all evaluation methods (see Guideline 3.1) should be presented individually, and differences and similarities in findings should be highlighted.

Guideline 5.2. Describe the impact of post data collection mitigation strategies on reducing nonresponse bias

While one aim of nonresponse bias analysis is to quantify the amount and nature of the bias introduced by nonresponse, an equally important aim is to identify possible mitigation strategies based on the results of the nonresponse bias analysis and apply them to the dataset. Ideally, these strategies can be tested on the current collection set (or subset thereof) but may have to be postponed to subsequent collections (see Best Practice 7). Once those strategies are implemented, it is then necessary to conduct a subsequent nonresponse bias analysis on the mitigated data, using the same variables as in the original analysis. The impact of mitigation strategies on nonresponse bias reduction for key survey items should be discussed. In addition, the extent of bias that remains after the mitigation strategies have been applied should also be presented.

Post data collection adjustments to mitigate nonresponse bias often involve a tradeoff between variance and bias. In addition to describing the impact on reducing bias, the discussion should also describe how the properties of the survey weights and the survey design effects may have changed.

Guideline 5.3. Describe nonresponse bias before and after adjustment across key stages and subgroups

The results of nonresponse bias analysis done after adjustment should be reported by each sampling phase and for each wave of data collection in longitudinal or panel studies as well as for key subgroups, including those for which geographic and demographic estimates are published (see Guideline 2.2 to 2.6). The change in nonresponse bias before and after adjustment

by stages, waves, and subgroups should be discussed to evaluate the impact of the adjustment strategies and to highlight the remaining bias that will need to be considered when using the data.

Best Practice 6: Summarize the major conclusions of the analyses

The report should provide a summary that includes high-level findings for the full data collection activity and for the key survey items, overall and by studied subgroup. If possible, the report should highlight post-data collection mitigation strategies that appear to be effective, as well as those that are less effective.

Guideline 6.1. Summarize the results of the nonresponse bias analysis, any post-data collection mitigation strategies, and the final assessment of potential nonresponse bias for the full data collection and for key indicators

All results of the nonresponse bias analysis should be summarized so that the reader has a good understanding of how the analyses were done and the magnitude and nature of bias that is not addressed through the mitigation process. This summary should not be merely an abbreviation of the findings, but a narrative that tells the story in a clear and concise manner.

When summarizing the results based on auxiliary data, information on the strength of the relationship between the auxiliary data and the survey data should be considered, as well as the quality of the auxiliary data. The summary should put greater emphasis on results that are obtained using strongly related auxiliary data as predictors or as benchmarks as opposed to findings via indirect assessments. When presenting results for specific outcome variables, the summary should include results by studied subgroups as well as for the full data collection, since nonresponse bias for each key survey item can vary by subgroup.

Assessments of post-data collection mitigation strategies based on the nonresponse bias analysis may vary by the key survey items. The summary should emphasize results from implemented strategies that have effects on nonresponse bias. Analyses and mitigation activities are often done in cycles which has the advantage of showing the effect that each mitigation strategy has on reducing observed bias. The nonresponse analysis completed after the last cycle of mitigation will be of most relevance to data users in their determination of whether the data are fit for their purposes and if any residual bias could impact their interpretation of the results of their own analyses.

Guideline 6.2. Discuss the implications and potential causes of contradictory results

Best Practice 3 advocates employing a variety of methods to assess nonresponse bias to obtain a more complete evaluation of bias. When more than one analysis method is applied to the same outcome variable, similarities and differences in the indicators of bias should be discussed, emphasizing the (cumulative) abundance of evidence for or against the presence of nonresponse bias. Contradictory results for the same outcome variable could be due to a variety of factors, including differences in the strength of the relationships between different sources of auxiliary data and survey estimates, auxiliary data shown to be of varying quality, or inappropriate assumptions in the assessment method, among others. Evaluate the relative strengths and weaknesses of the methods used with respect to the studied outcome variable. If possible, the summary should emphasize results obtained with the stronger methods, without completely discounting the others.

Guideline 6.3. Discuss potential implications as they relate to stated objectives of the data collection

The result of the nonresponse bias analysis and the impact of any mitigation strategies conducted should be discussed in terms of the core objectives of the data collection. The results should be discussed in relation to the key items in terms of whether the level and type of residual bias affect the usefulness of the data for the main objectives of the collection. Overwhelming evidence of nonresponse bias in a key item in one or more subgroups can affect the ability to use the data to meet the core data collection objectives. To the extent possible, provide guidance to the potential data user regarding strengths and limitations of the data for the objectives for which the data were collected.

Best Practice 7: Discuss recommendations for data collection methods and adjustment strategies to mitigate nonresponse bias in future waves or iterations of longitudinal and cross-sectional data collection

A thorough nonresponse bias analysis that follows the previous guidelines for reporting should provide useful insight into the likely sources of nonresponse bias. In recurring surveys, whether cross-sectional or longitudinal, this information can be used to improve future data collection. These improvements may either reduce response bias during data collection or improve the ability to reduce bias after data collection. Recommendations for how the results of the current evaluation may be used to mitigate nonresponse in the future should be offered and discussed.

Guideline 7.1. Discuss recommendations for modifications to sample design, questionnaires, or data collection strategies that may reduce the bias in collected data

The evaluation of nonresponse bias and mitigation strategies may identify potential sources of nonresponse bias. These insights can be used to suggest modifications to sampling, such as respecifying strata or oversampling groups with lower response propensities, or to future data collection strategies by including adaptive or responsive design. They may inform modifications to the questionnaire, such as relocating or removing items that result in high rates of breakoffs. The results of experiments embedded in previous waves may suggest optimal follow-up strategies or incentive structures. Based on the results described in Best Practice 5, the report should provide recommendations for further reducing nonresponse bias in collected data. Those recommendations may include specific ideas for research to better understand and address the impact of data collection strategies on nonresponse bias. Recommendations may also include design changes for sampling stages, waves, or key subgroups to increase sample sizes and enhance the power of post adjustment methods.

Guideline 7.2. Discuss recommendations that may improve post-data collection adjustment strategies

The report should offer and discuss recommendations for improving nonresponse mitigation and adjustment strategies following future data collections, given the results of the post-data collection adjustment strategies – such as statistical imputation, poststratification, raking, adjustment cell weighting and weight trimming – that were used. Recommendations may also include questionnaire modifications to better align constructs with existing benchmarks or collect new data that may permit more complex or refined weighting adjustments and imputations. If the availability of resources limited the scope of the present nonresponse bias

analysis or post-data collection adjustment strategies, recommendations could also include a description of the desirable approaches that could not be evaluated. Recommendations should be discussed in terms of the tradeoffs between variance and nonresponse bias. The report should provide recommendations, if any, for the use of decision criteria for managing these tradeoffs. The recommendations may also include specific ideas for research to better understand and address post data collection adjustment strategies to mitigate nonresponse bias.

REFERENCES

American Association for Public Opinion Research (AAPOR). (2023). Standard definitions: Final dispositions of case codes and outcome rates for surveys, 10th edition. https://aapor.org/wp-content/uploads/2023/05/Standards-Definitions-10th-edition.pdf

Andridge, R., & Thompson, K. J. (2015). Using the fraction of missing information to identify auxiliary variables for imputation procedures via proxy pattern-mixture models. International Statistical Review, 83(3), 472-492.

Barge, S., & Gehlbach, H. (2012). Using the theory of satisficing to evaluate the quality of survey data. Research in Higher Education, 53(2), 182-200.

Biemer, P. P. (2010). Total survey error: design, implementation, and evaluation. Public Opinion Quarterly, 74(5), 817–848.

Brick, J. M. (2013). Unit nonresponse and weighting adjustments: A critical review. Journal of Official Statistics, 29(3), 329-353.

Earp, M., Mitchell, M., McCarthy, J., & Kreuter, F. (2014). Modeling nonresponse in establishment surveys: Using an ensemble tree model to create nonresponse propensity scores and detect potential bias in an agricultural survey. Journal of Official Statistics, 30(4), 701-719.

Earp, M., Toth, D., Phipps, P., & Oslund, C. (2016). Identifying establishment characteristics of potential nonresponse bias. Bureau of Labor Statistics: Washington, DC. https://www.bls.gov/osmr/research-papers/2016/pdf/st160020.pdf

Earp, M., Toth, D., Phipps, P., & Oslund, C. (2018). Assessing nonresponse in a longitudinal establishment survey using regression trees. Journal of Official Statistics (JOS), 34(2).

Federal Committee on Statistical Methodology (FCSM). (2020). A framework for data quality. FCSM 20-04. Federal Committee on Statistical Methodology: Washington, DC. https://www.fcsm.gov/assets/files/docs/FCSM.20.04_A_Framework_for_Data_Quality.pdf

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. Public Opinion Quarterly, 70(5), 646-675.

Groves, R. M. (2011). Three eras of survey research. Public Opinion Quarterly, 75(5), 861-871.

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: a meta-analysis. Public Opinion Quarterly, 72(2), 167-189.

Heerwegh, D., Abts, K., & Loosveldt, G. (2007, January). Minimizing survey refusal and noncontact rates: do our efforts pay off?. In Survey Research Methods (Vol. 1, No. 1, pp. 3-10).

Hendra, R., & Hill, A. (2019). Rethinking response rates: new evidence of little relationship between survey response rates and nonresponse bias. Evaluation Review, 43(5), 307-330.

Iwig, W., Berning, M., Marck, P., & Prell, M. (2013). Data quality assessment tool for administrative data. Federal Committee on Statistical Methodology: Washington, D.C. https://www.fcsm.gov/assets/files/docs/DataQualityAssessmentTool.pdf

Johnson, T. P., & Smith, T. W. (2017). Big data and survey research: Supplement or substitute?. In: Thakuriah, P., Tilahun, N., Zellner, M. (eds), Seeing cities through big data. Springer Geography. Springer, Cham.

Kreuter, F. (Ed.). (2013). Improving surveys with paradata: Analytic uses of process information (Vol. 581). John Wiley & Sons.

Kreuter, F., Olson, K., Wagner, J., Yan, T., Ezzati-Rice, T. M., Casas-Cordero, C., Lemay, M., Peytchev, A. & Raghunathan, T. E. (2010). Using proxy measures and other correlates of survey outcomes to adjust for non-response: examples from multiple surveys. Journal of the Royal Statistical Society: Series A (Statistics in Society), 173(2), 389-407.

Krueger, B. S., & West, B. T. (2014). Assessing the potential of paradata and other auxiliary data for nonresponse adjustments. Public Opinion Quarterly, 78(4), 795-831.

Lepkowski, J. M., Mosher, W. D., Groves, R. M., West, B. T., Wagner, J., & Gu, H. (2013). Responsive design, weighting, and variance estimation in the 2006-2010 National Survey of Family Growth. National Survey for Health Statistics. Vital and Health Statistics, 2(158), 1-52.

Little, R. J., & Vartivarian, S. (2005). Does weighting for nonresponse increase the variance of survey means?. Survey Methodology, 31(2), 161-168.

Lohr, S. L. (2004). Small area and longitudinal estimation using information from multiple surveys. Statistical Policy, 77.

Lundquist, P., & Särndal, C-E. (2013). Aspects of responsive design with applications to the Swedish Living Conditions Survey. Journal of Official Statistics, 29(4), 557-582.

Maitland, A., Casas-Cordero, C., & Kreuter, F. (2009, August). An evaluation of nonresponse bias using paradata from a health survey. In Proceedings of the Section on Survey Research Methods of the American Statistical Association (pp. 370-78).

Massey, D. S., & Tourangeau, R. (2013). Where do we go from here? Nonresponse and social measurement. The ANNALS of the American Academy of Political and Social Science, 645(1), 222-236.

McConville, K. S., & Toth, D. (2019). Automated selection of post-strata using a model-assisted regression tree estimator. Scandinavian Journal of Statistics, 46(2), 389-413.

Merkle, D. M., & Edelman, M. (2009). An experiment on improving response rates and its unintended impact on survey error. Survey Practice, 2(3), 2964.

Miller, P., Fakhouri, T. H., Earp, M., Piscopo, K. D, Frenk, S. M. Christopher, E., & Madans, J. (2020). A systematic review of nonresponse bias studies in federally sponsored surveys. FCSM 20-02. Federal Committee on Statistical Methodology. March 2020. https://www.fcsm.gov/assets/files/docs/A_Systematic_Review_of_Nonresponse_Bias_Studies_Federally_Sponsored_SurveysFCSM_20_02_032920.pdf

Montaquila, J. M., Brick, J. M., Williams, D., Kim, K., & Han, D. (2013). A study of two-phase mail survey data collection methods. Journal of Survey Statistics and Methodology, 1(1), 66-87.

Office of Management and Budget. (2006). Standards and guidelines for statistical surveys. https://www.whitehouse.gov/wp-content/uploads/2021/04/standards_stat_surveys.pdf

Olson, K. (2013). Paradata for nonresponse adjustment. The Annals of the American Academy of Political and Social Science, 645(1), 142-170.

Peytcheva, E., & Groves, R. M. (2009). Using variation in response rates of demographic subgroups as evidence of nonresponse bias in survey estimates. Journal of Official Statistics, 25(2), 193-201.

Phipps, P., & Toth, D. (2012). Analyzing establishment nonresponse using an interpretable regression tree model with linked administrative data. The Annals of Applied Statistics, 772-794.

Prell, M., Chapman, C., Adeshiyan, S., Fixler, D., Garlin, T., Mirel, L., & Phipps, P. (2019). Transparent reporting for integrated data quality: Practices of seven federal statistical agencies. FCSM 19-01. Federal Committee on Statistical Methodology. September 2019. https://www.fcsm.gov/assets/files/docs/Transparent_Reporting_FCSM_19_01_092719.pdf

Rao, J. N. K. (2003), Small area estimation, New York: Wiley.

Redman, T. C. (2001). Data quality: the field guide. Boston: Digital Press.

Särndal, C.-E., & Lundström, S. (2008). Assessing auxiliary vectors for control of nonresponse bias in the calibration estimator. Journal of Official Statistics, 24(2), 167-191.

Schouten, B., Bethlehem, J., Buellens, K., Kleven, Ø., Loosveldt, G., Luiten, A., Rutar, K., Shlomo, N., & Skinner, C. (2012). Evaluating, comparing, monitoring, and improving representativeness of survey response through R-indicators and partial R-indicators. International Survey Review, 80(3), 382–399.

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

Wagner, J., West, B. T., Kirgis, N., Lepkowski, J. M., Axinn, W. G., & Ndiaye, S. K. (2012). Use of paradata in a responsive design framework to manage a field data collection. Journal of Official Statistics, 28(4), 477-499.

West, B. T. (2013). An examination of the quality and utility of interviewer observations in the National Survey of Family Growth. Journal of the Royal Statistical Society: Series A (Statistics in Society), 176(1), 211-225.

West, B. T. & Kreuter, F. (2015). A practical technique for improving the accuracy of interviewer observations of respondent characteristics. Field Methods, 27(2), 144-162.

West, B. T. & Little, R. J. A. (2013). Non-response adjustment of survey estimates based on auxiliary variables subject to error. Journal of the Royal Statistical Society: Series C (Applied Statistics), 62(2), 213-231.

APPENDIX A: AUXILIARY DATA SOURCE TYPES

Frame Data

One of the most common types of auxiliary data is sample frame data. The richness of the frame data — that is, the extent and nature of the data on all eligible sample units — tends to vary across establishment and household surveys (Brick, 2013; Groves, 2006). Establishment surveys like the Jobs Openings and Labor Turnover and the Occupational Requirements Surveys rely on rich sampling frames like the Quarterly Census of Employment and Wages (QCEW) that provide auxiliary data on establishment size, age, sector, location, employment rate, and wage data (Phipps & Toth, 2012; Earp et al., 2016). Household surveys like the Current Population Survey and the Current Expenditure Survey rely on less rich and at times less current sampling frames like the Master Address File. Most agriculture surveys use the Census of Agriculture as a frame, which provides an extensive list of farms, containing a rich source of household and establishment-level variables for farm operations (Earp et al., 2014).

Establishment surveys may have access to richer frames than household surveys in terms of auxiliary data and may also include more relatively current census data. For example, the QCEW collects establishment data every quarter and the Economic Census is conducted every five years; in contrast, the Decennial Census collects household data every ten years (Brick, 2013). As the time between sample selection and data collection increases, the comparability of the frame variables and survey items decreases. Furthermore, sampling unit compositions and classification can change over time. For example, a company might grow or shrink due to acquisitions or divestitures; an establishment could change industries when a different business moves into the same location.

Administrative Records

Administrative data differ from survey data, even when both collect similar variables. Survey data are collected for statistical purposes, allowing for instrument design tailored to specific outcomes. Administrative data are collected to facilitate the operation of programs and reflect the requirements of those programs. Consequently, there can be differences in the collection units, which in turn can affect comparisons. It should not be assumed that administrative data are not subject to measurement error. Rather survey and administrative data are subject to different sources of measurement error related to how the data are collected and who is providing the information. As with other sources of auxiliary data, these potential differences must be considered when evaluating whether differences are due to survey nonresponse bias.

Paradata

According to Kreuter (2013) and Olson (2013), paradata are data that are actively or passively collected during the data collection process, can be collected in face to face, telephone, and web surveys, and including everything from contact history information and interviewer observations to potential mail tracing and respondent keystroke entries. An example of paradata use in a nonresponse bias analysis is the evaluation by Maitland et al. in the National Health Interview Survey (Maitland, Casas-Cordero, & Kreuter, 2009) This evaluation used data about what types of efforts were made to contact the household and data on household cooperation (mostly consisting of reasons households gave for not wanting to participate) to improve previous models of non-response bias in the survey.

Benchmarks

Benchmarks are aggregate estimates (totals) that provide independently obtained estimates of the survey total. Differences between benchmarks and (appropriately weighted) respondent totals are often provided as estimates of nonresponse bias.

Gold standard benchmarks should be valid measures of the construct, be defined similarly, have comparable coverage, and have high reliability in terms of sampling and nonsampling errors. Any possible differences in measurement, coverage, and reliability between the survey totals and corresponding benchmarks should be discussed in the nonresponse bias analysis before providing comparisons.

APPENDIX B: AUXILIARY DATA QUALITY METRICS

Missingness

Considerations of missingness in auxiliary data might be different depending on whether the auxiliary data are survey, administrative, paradata, or other types of data. For example, if paradata or contact history instrument (CHI) data are used, it is important to understand the nature and extent of missing data. Missingness in the auxiliary data may be correlated with missingness in the survey data and may limit the accuracy of the findings.

Measurement Error

Auxiliary data typically comes from another source of data source and may have slightly different measurement properties as a result. According to Biemer et al. (2013), there are three main sources of measurement error to consider: 1) measuring the wrong variable; 2) incorrect or biased measures; and 3) measurement variability due to the instrument. When assessing the appropriateness of auxiliary data, consider the appropriateness of the way the variable is conceptualized, any known measurement biases, and any known measurement variability.

Coverage

In the survey design literature, auxiliary data coverage describes overlap between the target population and the auxiliary data source. Known differences of coverage in auxiliary data should be noted when conducting and reporting nonresponse bias analyses.

On occasion, an excellent source of auxiliary data will overlap with only part of the target population. In this case, the nonresponse bias analysis is valid for the overlap segment, but the analyses should not be extrapolated to the entire survey. For example, the target population could be the entire United States, but the auxiliary data could contain data collected solely in the

Midwest. In this case, the nonresponse bias analysis would be valid for the Midwest (assuming the survey data are representative of the Midwest), but not for the Northeast, South, or West census regions.

Timeliness

For nonresponse bias analyses, the terms "timeliness" and "age" are interchangeable and refer to the difference between the survey's reference period for data collection and the reference period for the auxiliary data's reference period. Both should be provided in the nonresponse bias analysis report.

The age of the auxiliary data alone may not be a concern if some assessment has been done to ensure that data are still up to date in terms of the survey data collection period. However, older data may indicate a higher likelihood of divergence from the survey data. It is important for the analyst to know when the data were last collected and updated or checked for accuracy and to match the reference periods for data collection especially in cases where population fluctuations are known to happen at given intervals (e.g., households moving during the summer, or numbers of children enrolled in a school at the end of each school term).