A Systematic Review of Nonresponse Bias Studies in Federally Sponsored Surveys

FCSM-20-02

Prepared by

Peter Miller, Tala H. Fakhouri, Morgan Earp, Kathryn Downey Piscopo, Steven M. Frenk, Elise Christopher, and Jennifer Madans

under the auspices of
Federal Committee on Statistical Methodology

Statistical and Science Policy Office
Office of Information and Regulatory Affairs
Office of Management and Budget

March 2020

Recommended citation: Peter Miller, Tala H. Fakhouri, Morgan Earp, Kathryn Downey Piscopo, Steven M. Frenk, Elise Christopher, and Jennifer Madans. 2020. *A Systematic Review of Nonresponse Bias Studies in Federally Sponsored Surveys.* FCSM 20-02. Federal Committee on Statistical Methodology. March 2020.

THE FEDERAL COMMITTEE ON STATISTICAL METHODOLOGY

MEMBERS

Rochelle (Shelly) Martinez (Co-Chair)
Office of Management and Budget

Stephen Blumberg
National Center for Health Statistics

Chris Chapman
National Center for Education Statistics

Jennifer Edgar
Bureau of Labor Statistics

John Eltinge
U.S. Census Bureau

John Finamore
National Center for Science and Engineering Statistics

Dennis Fixler
Bureau of Economic Analysis

Michael Hawes
U.S. Census Bureau

Jennifer Madans
National Center for Health Statistics

Wendy Martinez
Bureau of Labor Statistics

Jaki McCarthy
National Agricultural Statistics Service

Jennifer Nielsen (Secretary)
National Center for Education Statistics

Jennifer Ortman
U.S. Census Bureau

Anne Parker
Internal Revenue Service

Jennifer Parker
National Center for Health Statistics

Polly Phipps
Bureau of Labor Statistics

Mark Prell
Economic Research Service

Joseph Schafer
U.S. Census Bureau

Rolf Schmitt
Bureau of Transportation Statistics

Marilyn Seastrom
National Center for Education Statistics

Joy Sharp
Energy Information Administration

Bob Siviniski
Office of Management and Budget

Bruce Taylor (Webmaster)
National Center for Education Statistics

G. David Williamson
Centers for Disease Control and Prevention

Linda Young
National Agricultural Statistics Service

MEMBERS OF THE NONRESPONSE BIAS ANALYSIS SUBCOMMITTEE

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

Tala H. Fakhouri (Co-Chair)
National Center for Health Statistics

Peter Miller
U.S. Census Bureau

Morgan Earp
Bureau of Labor Statistics

Kathryn Downey Piscopo
Substance Abuse and Mental Health Services Administration

Steven M. Frenk
National Center for Health Statistics

Elise Christopher
National Center for Education Statistics

ABSTRACT

Background: In 2006, the Office of Management and Budget (OMB) published *Standards and Guidelines for Statistical Surveys* mandating that all federal surveys with a unit response rate of less than 80% conduct an analysis of nonresponse bias (NRB). Since 2006, federal surveys have increased activities involving NRB analyses; however, it is unclear what methods have been used to assess NRB or whether mitigating strategies reduced bias.

Objective: This paper provides the first systematic review of NRB studies involving federal surveys since the release of the 2006 OMB *Standards and Guidelines for Statistical Surveys*. The objective of this systematic review was to collect NRB studies involving federal surveys and summarize the characteristics of the surveys examined, the NRB analysis methods used, and the assessment of NRB for each.

Methods: NRB reports involving federal surveys were identified via searches on PubMed, Google Scholar, Current Index to Statistics (CIS), Joint Statistical Meetings (JSM) proceedings, and through an open call to federal statistical agencies and associated professional organizations. The search yielded 425 documents. After inclusion criteria were applied, 165 eligible studies were identified. Through systematic coding, reviewers extracted information on the general characteristics of these studies (e.g., survey type and mode), the types of NRB assessment method used, the target of NRB analyses (sample composition, survey estimates, or both), and whether post-survey adjustments reduced bias. The reported NRB methods were grouped into four categories: benchmarking; comparisons to external data; studying variation within the respondent set; and comparing alternative post-survey adjustments.

Results: Eighty-nine of the 165 eligible studies were establishment surveys and 76 were household surveys. About 40% of the studies were conducted shortly after the release of the 2006 guidance. Comparisons of survey estimates to external data was the most commonly used NRB assessment method for establishment surveys (89.9%); whereas, studying variations within the respondent set was the mostly commonly used method in household surveys (65.8%). A majority of studies reported bias in some variables prior to weight adjustment but noted that bias was reduced in at least one of the variables after weighting (85.6%).

Conclusions: The types of NRB assessment methods used differed by survey type and mode of data collection. Most studies reported a reduction in bias after weighting. However, in most instances, a reduction in nonresponse bias was declared, but not explicitly documented, in the reports.

BACKGROUND

Over the last two decades, survey response rates have been steadily falling, with more accelerated declines reported in recent years (Czajka and Beyler 2016). Survey response rate is a valuable data quality measure and the most widely used indicator of survey quality. A high response rate increases the likelihood that the survey accurately represents the target population. However, a lower response rate is not always associated with higher levels of nonresponse bias (NRB), and the levels of NRB can differ for different estimates in the same survey.

In 2006, the Office of Management and Budget (OMB) published *Standards and Guidelines for Statistical Surveys* (Guidelines 1.3.3, 1.3.4, and 3.2.9) encouraging federal statistical agencies to assess the impact of declining response rates on the quality of official estimates if survey response rates dropped below 80% (Office of Management Budget 2006).

Prior to the 2006 guidance, OMB sponsored educational efforts to familiarize agency researchers with the threats posed by nonresponse bias and potential methods for detecting and reducing potential bias (Groves and Brick 2005). In 2009, the Federal Committee on Statistical Methodology (FCSM) sponsored a workshop on how to conduct NRB studies in household and establishment surveys (Federal Committee on Statistical Methodology 2009).

In 2010, OMB attempted to assess agency practices through a solicitation to agencies for NRB studies that had been conducted in response to the 2006 OMB *Standards and Guidelines for Statistical Surveys*. Some reports were collected, and a template was developed for categorizing and analyzing the reports; however, due to time constraints the project was not completed. In 2016, FCSM became engaged with assembling research on NRB across the federal statistical system, following some intra-agency efforts to examine the problem (Czajka and Beyler 2016). As a result, an FCSM subcommittee on survey NRB was created and charged with synthesizing and summarizing NRB assessment techniques and remedies.

In this report, we present findings from a systematic review of NRB studies involving federal surveys since the release of the 2006 OMB *Standards and Guidelines for Statistical Surveys*. The objective of the review was to summarize the characteristics of the surveys examined, the NRB analysis methods used, and the assessment of NRB for each. This information may be useful in developing guidelines and best practices for nonresponse bias analysis.

METHODS

Criteria for considering studies for this review

A collection of studies including peer-reviewed research papers, published book chapters, conference proceedings, published government reports or memoranda, and other grey literature materials of NRB studies involving federal surveys since the 2006 OMB guidance were considered for inclusion. Grey literature materials, as defined by Cochrane (Higgins and Green 2008), included unpublished reports produced by the government, and reports produced by academics and the survey industry in print and electronic formats not controlled by commercial publishers.

Search methods

This literature search built upon the 2010 OMB-sponsored effort to compile and classify NRB analyses conducted in connection with federally funded surveys. In 2010, an email was sent by the OMB Office of Information and Regulatory Affairs (OIRA) Statistical and Science Policy (SSP) to upper management staff contacts at the 13 existing federal statistical agencies. The email requested that establishment and household NRB reports or publications be sent to OMB for coding in a research study. All 13 agencies responded to the call. The request yielded 87 documents from 11 federal statistical agencies; two agencies had no nonresponse bias studies to report since their response rates were still above 80% on all their surveys. The 2010 effort was not completed due to time constraints. The compilation of documents submitted was not vetted for relevance or further analyzed.

For this systematic review, documents gathered in 2010 were scrutinized for relevance, and additional nonresponse bias studies were collected through a variety of methods. An online

literature search was conducted in March of 2017 using the Current Index to Statistics (CIS) (years: 2011–2015) and Joint Statistical Meetings (JSM) Proceedings (years: 2011 and 2013–2016). The JSM Proceedings encompasses papers presented at the JSM in addition to several other conferences sponsored by the American Statistical Association (ASA) or its sections. This search yielded 23 studies. In May 2017, another online literature search was conducted to identify peer reviewed published NRB studies. In that PubMed and Google Scholar search, the terms “nonresponse analysis”, “bias analysis”, and “survey nonresponse” were used for the years 2006 to 2017, identifying an additional 59 studies from various publications.

In November 2017, two solicitations were sent to the ASA and AAPORnet listservs asking for NRB studies of federal surveys. The goal of these solicitations was to identify grey literature reports and studies. The solicitation emailed are provided in **Supplemental File 1**. Emails were also sent to approximately 50 known authors of NRB studies. These solicitations gathered an additional 240 papers, some of which were duplicates.

In March of 2018, preliminary findings of this project were presented at the FCSM Research and Policy Conference. At the conference, the working group requested that members of the audience who had authored or co-authored NRB studies email the working group their reports. After this final solicitation, an additional 16 studies were gathered. Our comprehensive search yielded a total of 425 studies. The number of studies collected at each stage is shown in **Supplemental Table 1**.

Data extraction

The 425 reports identified by the search strategy were checked by two reviewers (Peter Miller and Kathryn Downey Piscopo) for duplicates. After duplicates were removed, the

remaining studies were divided and randomly assigned to the six members of the working group (Peter Miller, Tala Fakhouri, Morgan Earp, Kathryn Downey Piscopo, Elise Christopher, and Steven Frenk). The reports were independently reviewed to determine eligibility and culled to exclude ones that were: 1) conducted prior to 2006, which is when the OMB guidance was released, 2) non-federally funded surveys, 3) review papers or non-pertinent documents, and 4) reports that had not been cleared for public release or were in draft form. After these exclusion criteria were applied, 165 studies were deemed eligible for this analysis. It is possible that there are other relevant studies from this period that were not captured by the efforts described above.

Through systematic coding, using a standardized data extraction form shown in Supplemental File 2, the working group extracted information from the 165 eligible studies concerning the following attributes: 1) general characteristics, including agency sponsorship, response rates, type of survey, and mode of data collection; 2) types of NRB assessment method(s) used; 3) target of the NRB analyses (i.e., sample composition, survey estimates, or both); and (4) whether post-survey nonresponse adjustments were employed and if these adjustments appeared to reduce bias in final estimates.

Reliability in coding and reaching consensus

The NRB reports were very heterogeneous and finding the desired information in them was often difficult. The lack of a common format and standardization of reporting meant that coding often involved interpretation that could vary from coder to coder. For this reason, after completion of the initial coding, a second round of blinded reviews of the collated reports, were conducted by three reviewers (Peter Miller, Tala Fakhouri, and Morgan Earp). Morgan Earp blind coded all studies and was considered the gold standard reviewer. Peter Miller and Tala Fakhouri blind coded half of the studies each and their coding was compared to Morgan Earp's.

The inter-rater percent agreement for seven coded variables (i.e., agency sponsor of the analysis, survey type, survey mode, target of the NRB, bias reported prior to weight adjustment, bias reported after weight adjustment, and types of NRB method used) was calculated. The inter-rater percent agreement between Peter Miller and Morgan Earp, and between Tala Fakhouri and Morgan Earp, ranged from 80% to 94%, and from 65% and 90%, respectively. Differences in coding between the raters were adjudicated by consensus.

Data analysis

The types of NRB methods used to assess bias were coded into 14 distinct methodological approaches discussed by Groves and Brick, using the standardized data extraction form shown in **Supplemental File 2**. An "other" category was specified to capture all other methods. The reported methods were then grouped into four categories using the Groves and Brick typology (Groves and Brick 2005): (1) benchmarking, which corresponds to method 11 in the data extraction form; (2) making comparisons to external data, which corresponds to methods 21 to 24; (3) studying variation within the respondent set, which corresponds to methods 31 to 36; (4) and comparing alternative post survey adjustments, which corresponds to methods 41 to 44.

Surveys were grouped into either household or establishment surveys. Household surveys included those of the general population (e.g., adults, 18-70 years of age), special populations (e.g., former Department of Veterans Affairs (VA) hospital patients), or those reported as household surveys with no specific classification. Establishment surveys included farm, school, health care facility, other or unclassified facility establishment surveys. Analyses were performed using Stata version 13.1. We used descriptive statistics to present counts and

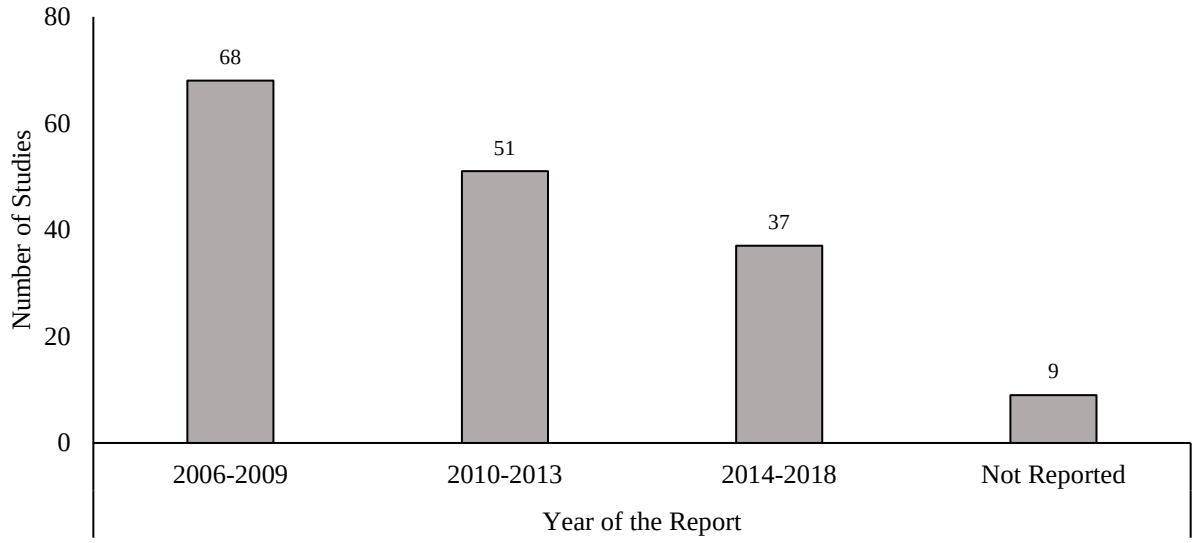
proportions by the variables of interest. Proportions (multiplied by 100 and expressed as percentages) are shown in the figures.

RESULTS

Characteristics of the studies

Characteristics of the studies coded are presented in Figures 1-10. The majority of studies were performed in the period immediately following the 2006 OMB *Standards and Guidelines for Statistical Surveys* (n=68), with smaller numbers completed during subsequent periods (**Figure 1**).

Figure 1. Number of nonresponse bias studies published by year from 2006 to 2018

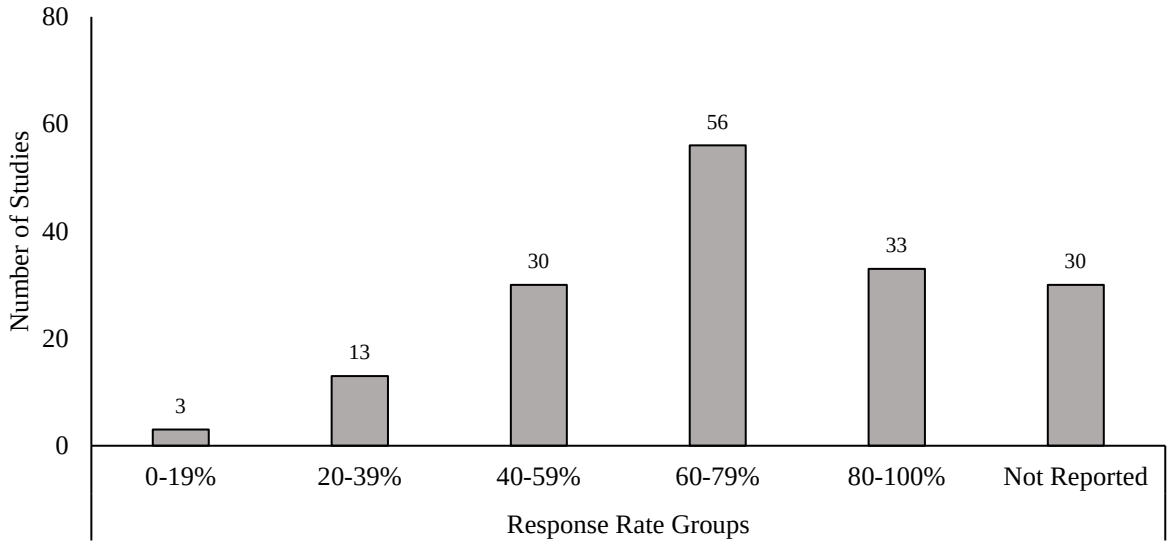


NOTES: 165 eligible studies were included in this analysis. "Not Reported" refers to reports with an unknown publication date but a known survey data collection date that met the eligibility criteria for this review.

The 2006 OMB *Standards and Guidelines for Statistical Surveys* encouraged federal statistical agencies to conduct analyses of NRB for any survey with a response rate below 80%. As shown in **Figure 2**, response rates were not reported in 30 studies. Of the 135 eligible studies that reported response rates, 102 of them (75.6%) had response rates below 80% (data not shown in the figure). The response rate reported for most of the studies ranged from 60% to 79%

(n=56). The 33 studies that had response rates of 80% or more were all conducted by the National Center for Education Statistics (NCES). NCES requires a NRB analysis to be conducted if the response rate falls below 85% (Seastrom and National Center for Education Statistics (NCES) 2002).

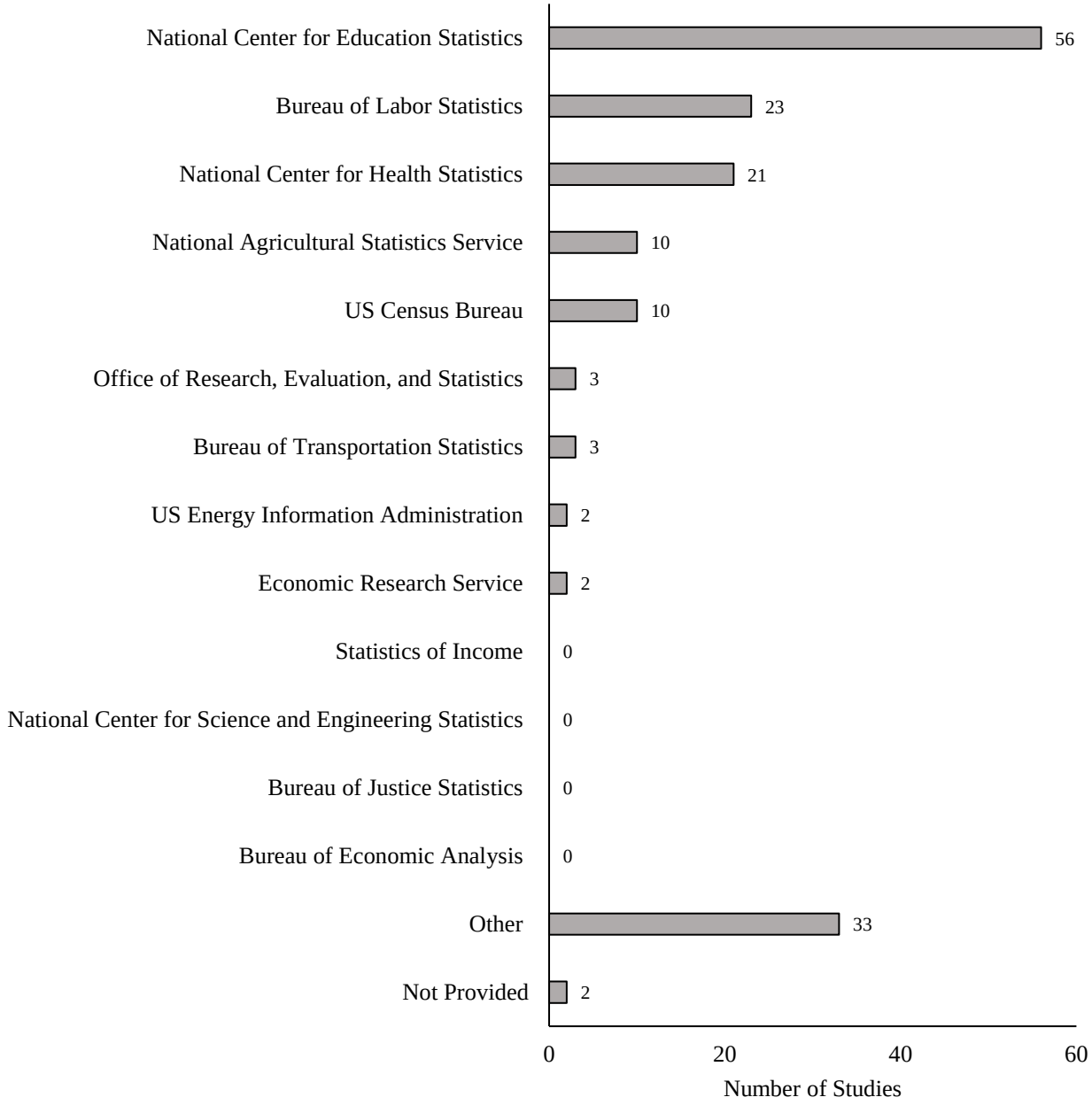
Figure 2. Number of nonresponse bias studies published by unit response rates



NOTE: 165 eligible studies were included in this analysis.

Figure 3 shows the distribution of studies by federal statistical agency sponsorship. The largest group of studies was sponsored by the NCES (n=56), followed by the Bureau of Labor Statistics (BLS) and the National Center for Health Statistics (NCHS), with 23 and 21 studies, respectively. Over 30 studies, grouped as "other", were federally sponsored by multiple statistical agencies or by non-statistical agencies (e.g., National Cancer Institute (NCI), VA, etc.). The large number of NCES-sponsored studies is mainly due to a large group of NRB studies done in connection with the National Assessment of Educational Progress (NAEP).

Figure 3. Number of nonresponse bias studies published by federal agency

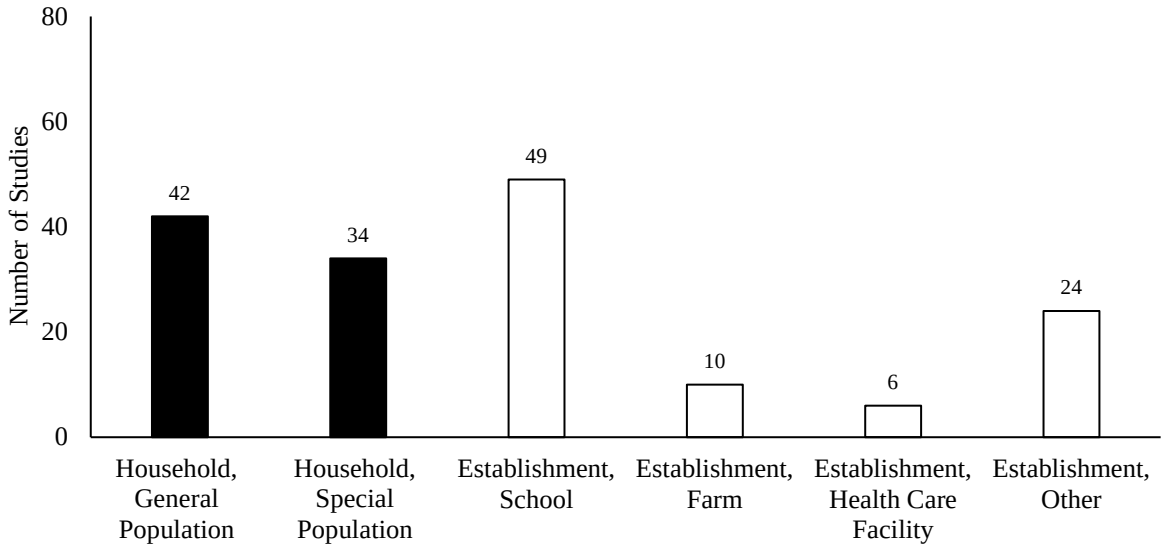


NOTES: 165 eligible studies were included in this analysis. The 13 federal statistical agencies are listed. The "other" category represents studies conducted by multiple statistical agencies or by non-statistical government agencies.

NRB studies were categorized by the type of survey – household or establishment. The majority of studies assessed focused on establishment surveys of one kind or another (i.e.,

schools, farms, healthcare facilities, or other types of establishments such as businesses or firms). The studies pertaining to household surveys included both general and special populations (i.e., surveys of adoptive parents, children, veterans, etc.). **Figure 4** shows the distribution of these studies in the sample. The large number of school establishment surveys is again attributable to the NAEP studies in the collection.

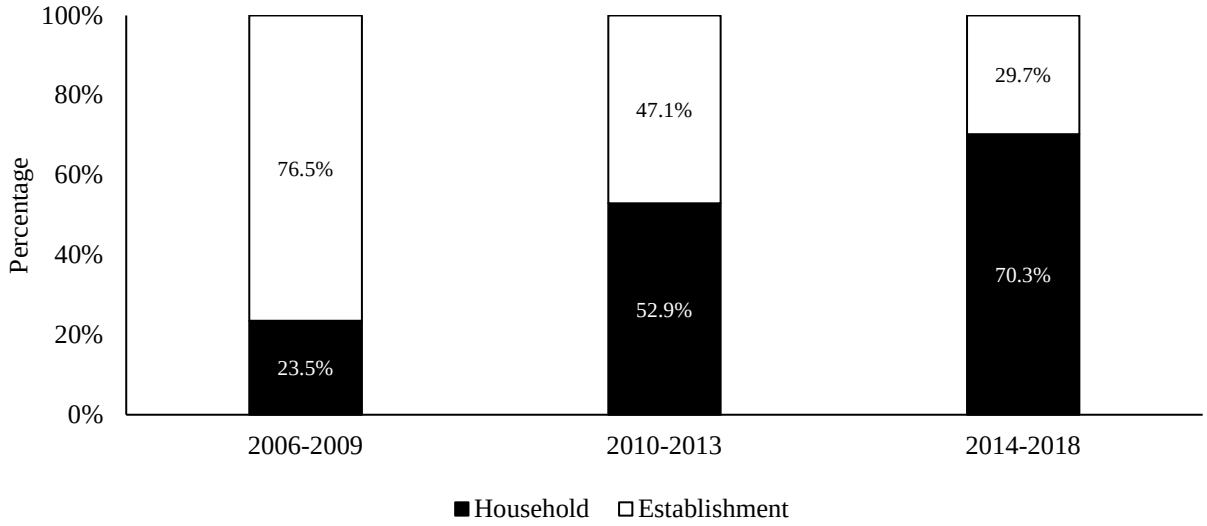
Figure 4. Number of studies published by survey type



NOTES: 165 eligible studies were included in this analysis. Black bars are household surveys and white bars are establishment surveys.

The proportion of studies from establishment surveys was higher in early years of the study period (proportions are multiplied by 100 and expressed as percentages in **Figure 5**). By contrast, the proportion of studies from household surveys grew over time. For example, from 2006 to 2009, 76.5% of NRB studies were from establishment surveys. By contrast, from 2014-2018, 70.3% of NRB studies were from houshold surveys.

Figure 5. Percentage of nonresponse bias studies published by survey type and by year from 2006 to 2018



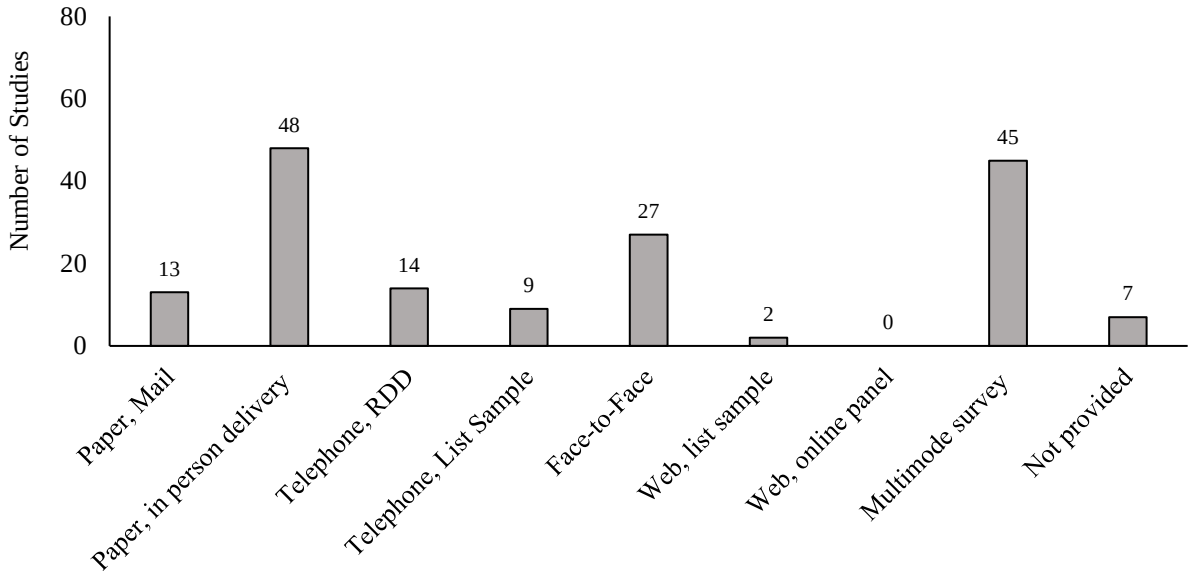
NOTES: 156 out of 165 eligible studies were included in this analysis. The 9 missing studies were those with an unknown publication date. Black bars represent the percentage of household surveys and white bars represent the percentage of establishment surveys by year.

Studies were also classified by the mode of data collection, the rationale being that the information available for NRB analysis would vary across modes. For example, the information available on a list sample for a telephone survey may be different from that obtained from a frame of an online panel. **Figure 6** shows the distribution of studies by mode. The largest categories were in-person delivery of paper questionnaire (i.e., commonly used in school-based surveys), multimode approaches, and face-to-face surveys.

The modes of data collection differed by survey type (**Figure 7**). Household surveys were more likely to utilize telephone (n=22), face-to-face (n=21), and multimode data collection methods (n=20). On the other hand, the majority of establishment surveys utilized mail surveys or paper questionnaires administered in schools (n=52), followed by multimode data collection methods (n=25). Multimode collections in establishment surveys are typically surveys conducted

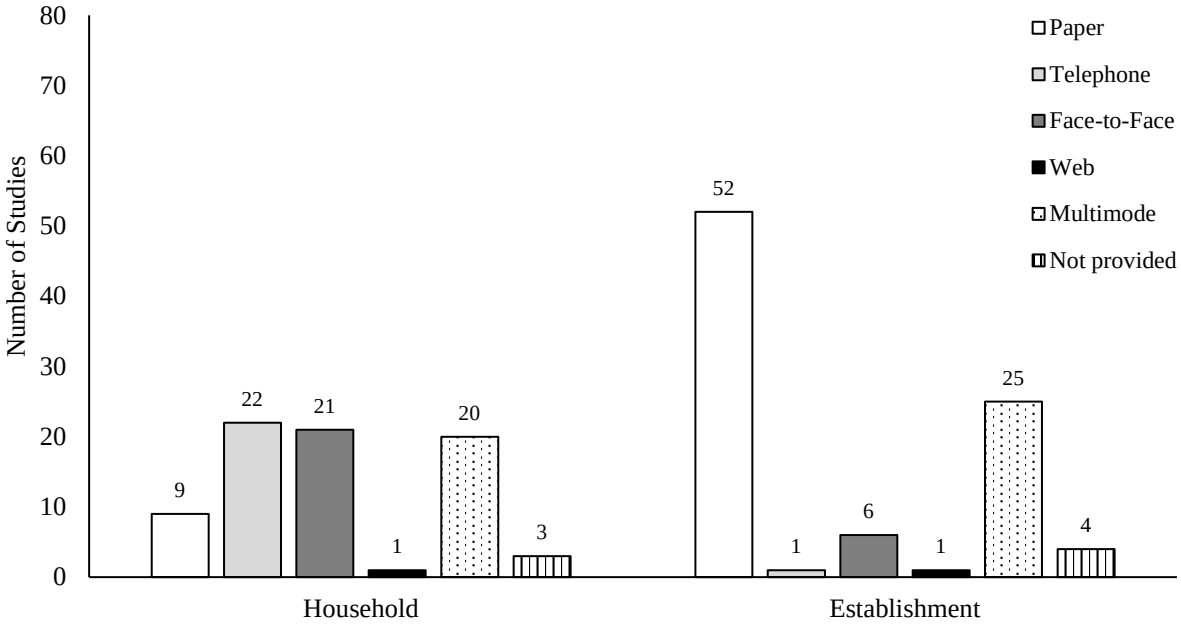
by mail with telephone follow-up to nonrespondents. Only one household survey and one establishment survey utilized web-based surveys as the sole mode of data collection.

Figure 6. Number of studies published by mode of data collection



NOTES: 165 eligible studies were included in this analysis. RDD is random digit dialing.

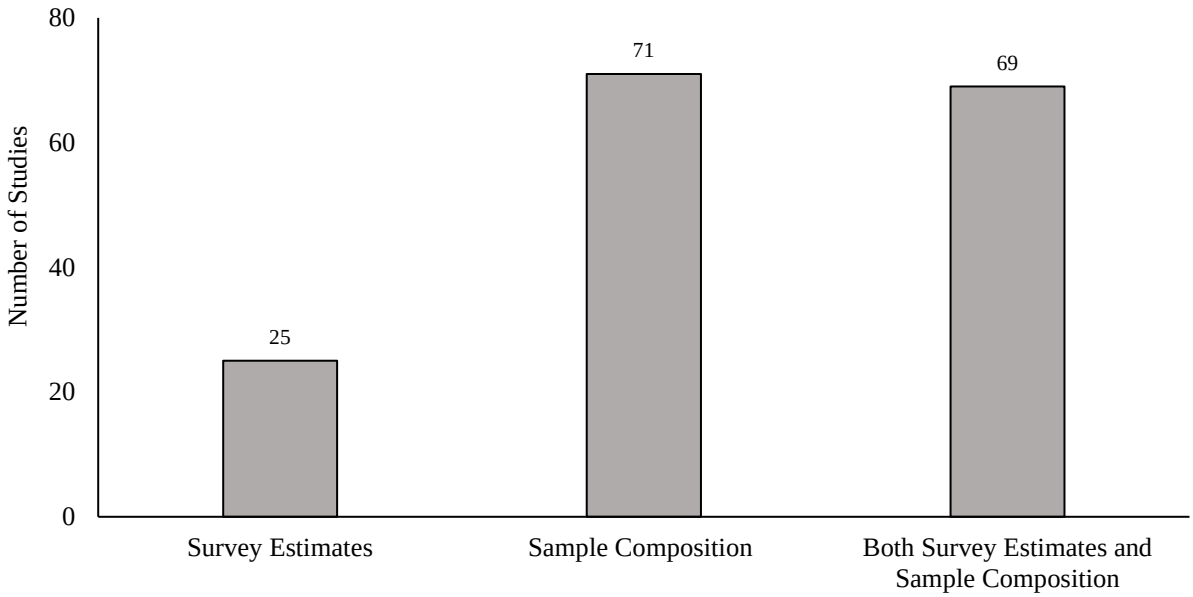
Figure 7. Number of studies published by mode of data collection and survey type



NOTE: 165 eligible studies were included in this analysis.

The focus or "target" of the NRB analysis in each study was coded to determine if the analysis looked for bias in sample characteristics (e.g., demographics or establishment characteristics), survey estimates (e.g., prevalence of health outcomes, proportion of U.S. adults currently enrolled in colleges, wages, etc.), or both. **Figure 8** shows that the largest groups coded were studies that looked for bias in sample characteristics (n=71) or both characteristics and estimates (n=69).

Figure 8. Number of studies published by target of analysis

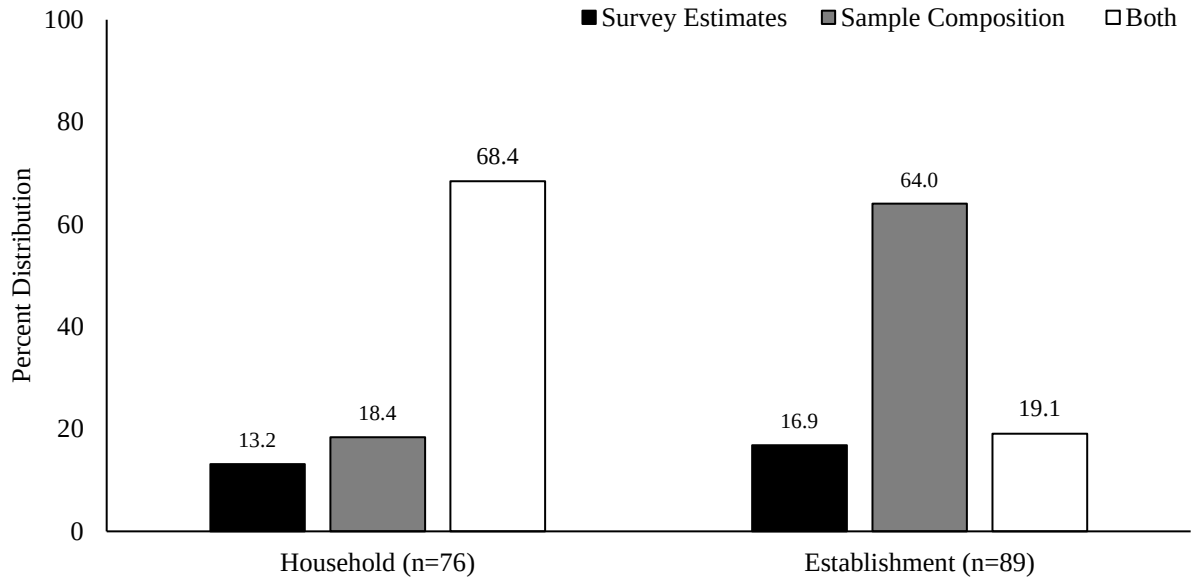


NOTE: 165 eligible studies were included in this analysis.

Next, we examined the target of the NRB analysis by survey type. The focus or "target" of the NRB analysis in each study differed by survey type **(Figure 9**). Household surveys were more likely to examine bias in both sample characteristics and survey estimates (68.4%); whereas, establishment surveys were more focused on studying bias in sample composition,

which is likely due to the availability of sample composition data on the frames of establishment surveys.

Figure 9. Percent distribution of nonresponse bias studies published by survey type and target of analysis

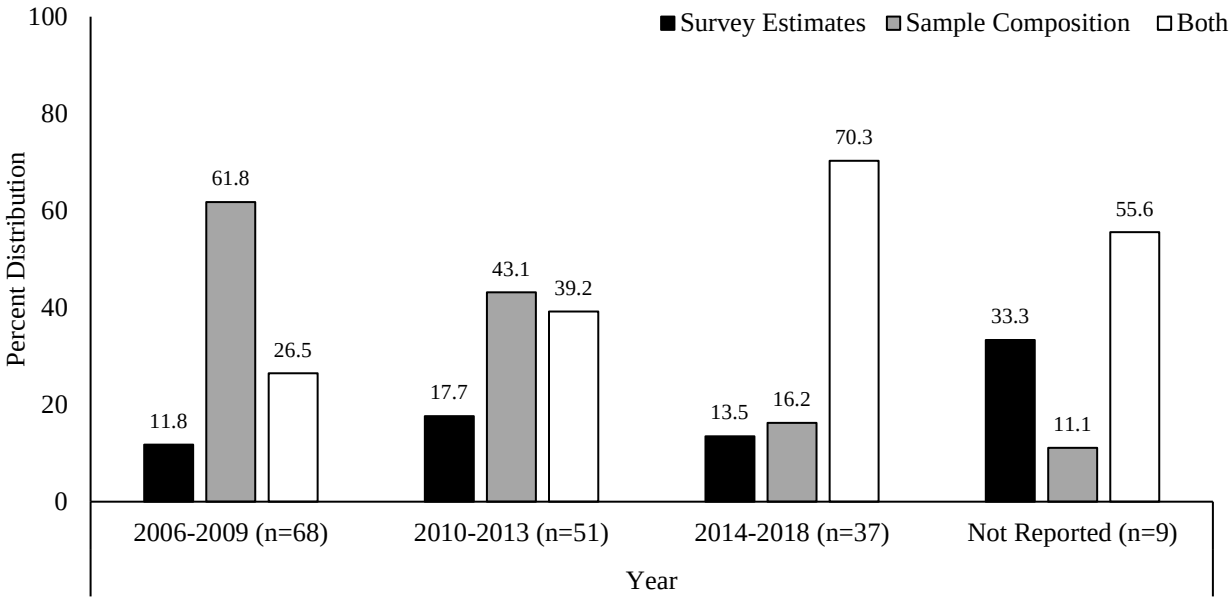


NOTE: 165 eligible studies were included in this analysis.

Next, we examined the target of the NRB analysis by year to investigate if the target of analysis changed over time. As illustrated in **Figure 10**, the proportion of studies that focused on bias in sample composition alone was higher in the earlier years. By contrast, the proportion of studies examining bias in both sample characteristics and survey estimates grew over time. These temporal trends may be explained, at least in part, by the change in the distribution of survey type (i.e., establishment versus household surveys) as was illustrated in Figure 5. Specifically, a larger proportion of the earlier studies came from establishment surveys and these surveys tend to focus on bias in sample composition. On the other hand, a larger proportion of studies in 2014

to 2018 came from household surveys, and these surveys tend to target both sample composition and survey estimates in NRB analyses.

Figure 6. Target of nonresponse bias analyses by year



NOTE: 165 eligible studies were included in this analysis.

Types of nonresponse bias methods used

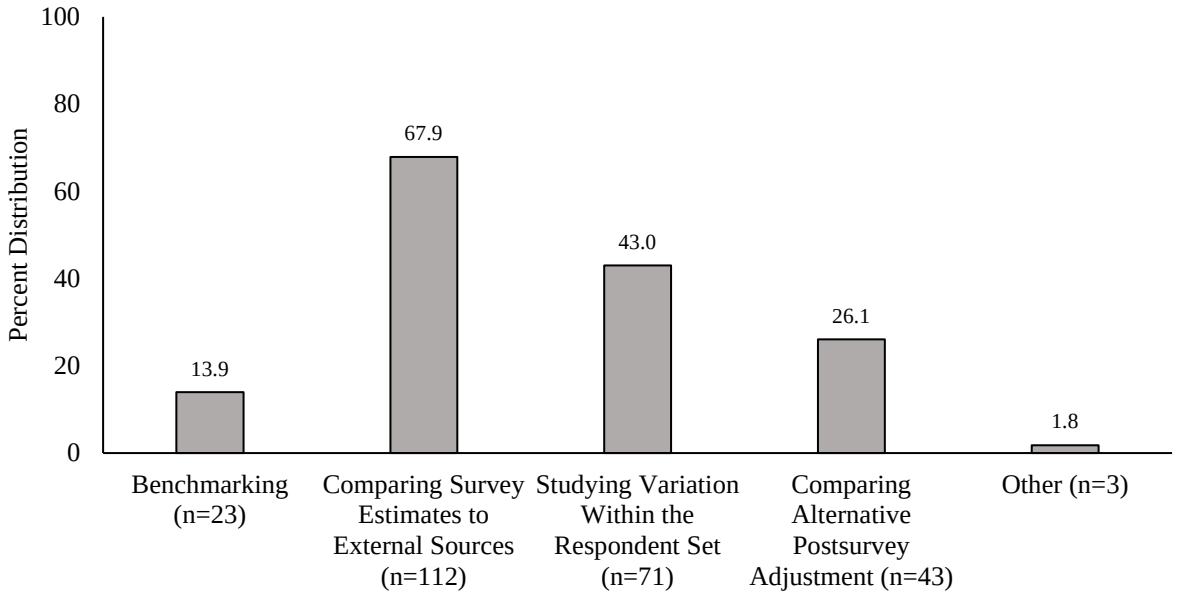
The methods used in NRB studies were coded and summarized according to the typology developed by Groves and Brick, and used by many federal statistical agency staff members during the period when the OMB guideline on conducting NRB analyses was introduced (Groves and Brick 2005). Groves and Brick noted that NRB studies could employ several approaches, namely: comparison to other, authoritative survey estimates (benchmarking); assessing survey estimates based on external information (e.g., information on the sampling frame); studying response variation within the responding sample (e.g., a level-of-effort analysis) and comparing the results of alternative weighting adjustments.

Each method has limitations. Assessing survey estimates based on external information typically concerns only a comparison of overall sample characteristics, on the assumption that the sample characteristics are correlated with survey estimates. Comparing alternative weighting schemes similarly relies on the strength of the correlations between demographic characteristics and survey estimates. On the other hand, studying variation within the respondent set does focus on survey estimates; for example, comparing values obtained from sample cases interviewed early and late in the field period, but it treats respondents who were harder to interview as proxies for nonrespondents. The assumption that hard-to-interview respondents are similar to nonrespondents has been shown to be unsupported in some cases (Lin and Schaeffer 1995, Teitler, Reichman et al. 2003). Finally, comparing survey estimates to measures available on the sampling frame is limited to those surveys that have suitable frame information.

Given the limitations of each method, employing ones with different limitations may allow analysts to "triangulate" on estimates of nonresponse bias (Campbell and Fiske 1959). There are challenges for interpretation and for developing summary bias measures; however, if the results differ by method. Groves and Brick, nonetheless, argued for employing multiple approaches.

Figure 11 displays the proportion of studies that employed each of these methods. The percentages do not add to 100% because some studies used multiple methods. Of the 165 studies included in this report, 37 household surveys and 29 establishment surveys reported using more than one method (date not shown in the figure). The two most commonly employed methods used external information to assess survey estimates and ones that examined variation in response within the responding sample, which partly depended on the type of survey (household versus establishment), and thus the type of frame data available.

Figure 7. Percent distribution of studies employing each of the major nonresponse bias methods



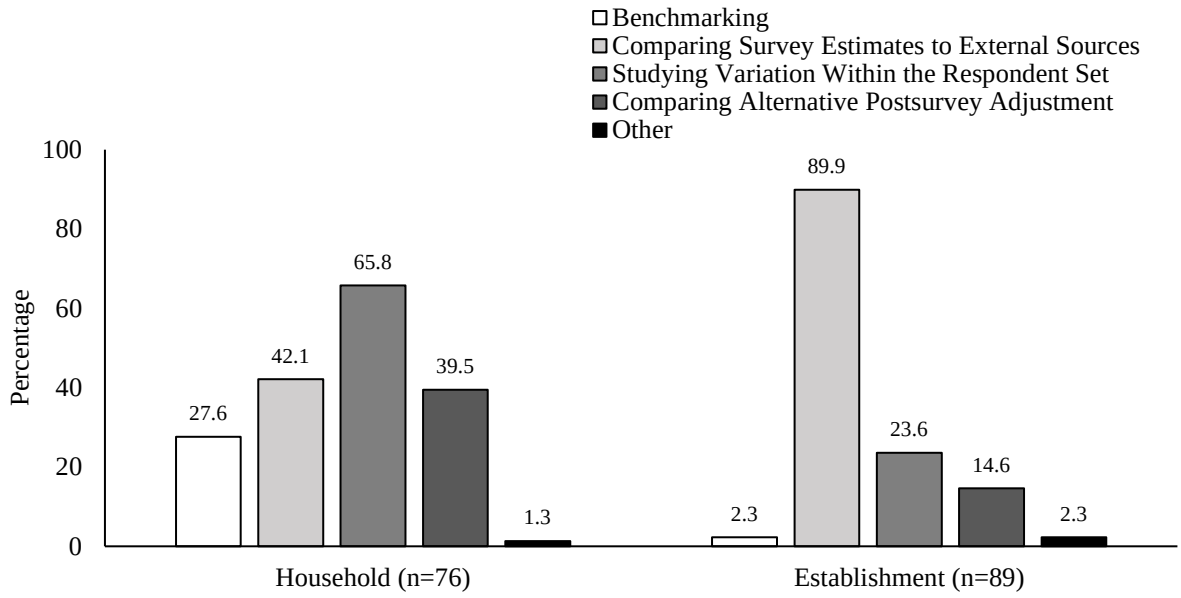
NOTES: 165 eligible studies were included in this analysis. Percentages do not add to 100% because some studies used more than one method.

The type of NRB analysis methods used differed by survey type (**Figure 12**). Household surveys were more likely to examine variation within the respondent set; whereas, establishment surveys were more likely to assess survey results against external data sources (e.g., the sampling frame). Establishment surveys tend to have richer frame information compared to household surveys where little is known about the household prior to data collection.

Next, we examined the pattern of NRB analysis methods used over time. Comparing survey estimates to external data sources was the most common method used at each time period, followed by studying response variation among respondents and comparing the results of alternative weighting approaches (**Figure 13**). While the patterns are similar, the use of the different methods appears to grow more equal over time. These temporal patterns can be explained in part by the changes in survey type over time, as was illustrated in Figure 5. A larger

proportion of studies originated from establishment surveys in the earlier years (i.e., 2006–2009), and establishment surveys are more likely to assess survey results against external data sources, as shown in Figure 12.

Figure 12. Nonresponse bias methods used by survey type

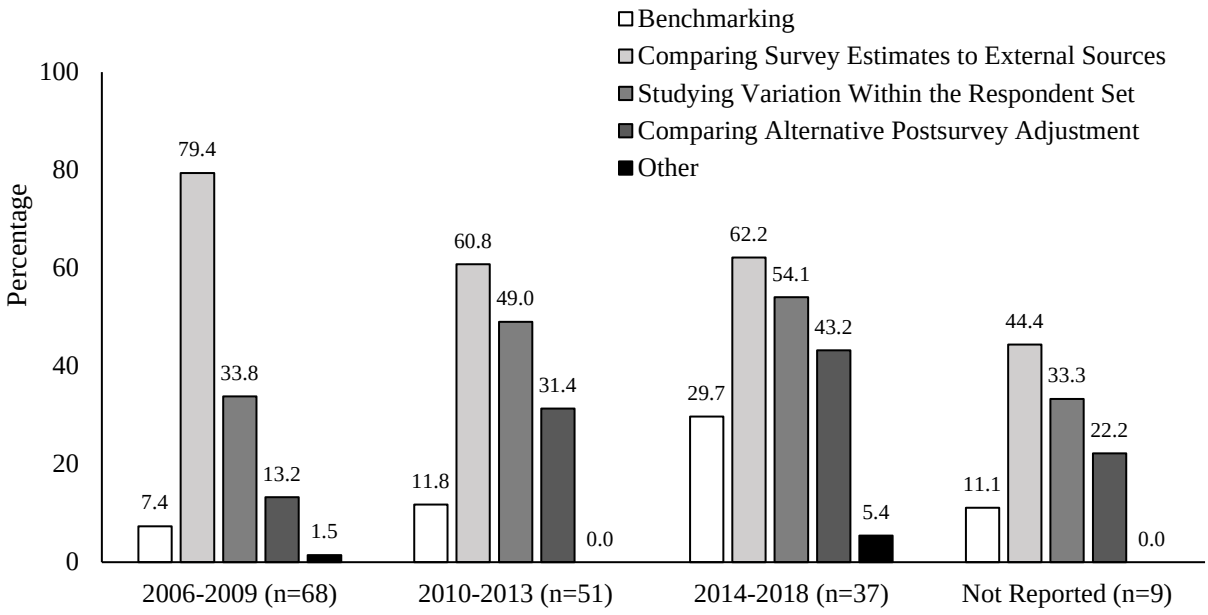


NOTES: 165 eligible studies were included in this analysis. Percentages do not add to 100% because some studies used more than one method.

The type of NRB analysis methods used also differed by data collection mode (**Figure 14**). For paper, web, and multimode surveys, the dominant method was assessing survey estimates against external data sources. The surveys using paper were those conducted in schools that had frame data available and the very few web surveys employed a list frame with auxiliary information. By contrast, telephone surveys and face-to-face surveys were more likely to examine response variation among respondents, which is in line with the typical lack of frame information in such studies. The most common analysis involving studying the variation within

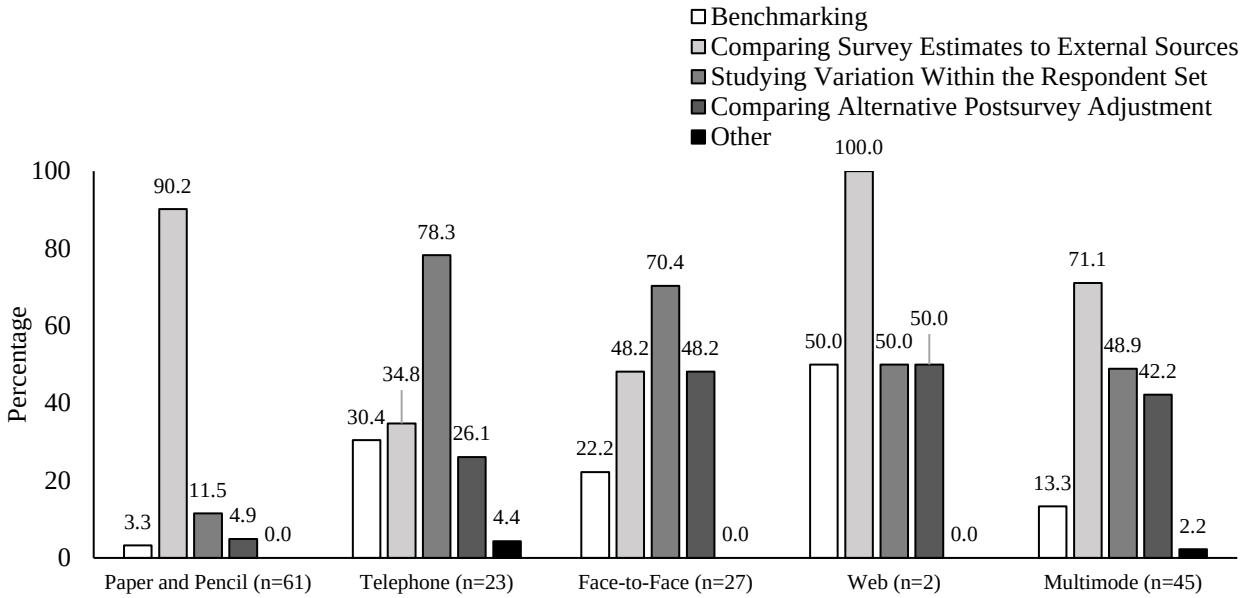
the respondent set, where survey estimates for early respondents were compared to late respondents.

Figure 13. Nonresponse bias methods used over time



NOTE: 165 eligible studies were included in this analysis. Percentages do not add to 100% because some studies used more than one method.

Figure 14. Nonresponse bias methods reported by mode



NOTES: 165 eligible studies were included in this analysis. Bars represent percentages. Percentages do not add to 100% because some studies used more than one method.

Finally, the impact of non-response weight adjustments on bias mitigation was examined. **Table 1** shows a cross-tabulation of NRB before and after weight adjustment. The majority of studies assessed reported bias in some variables prior to weighting and indicated that the bias was reduced in at least one variable by weighting. Because studies varied in their assessment and description of bias before and after weighting, no further comparisons or summaries were made.

Table 1. Report of nonresponse bias, before and after weight adjustment

	Bias Before Weighting			
	Yes (n=111)	No (n=3)	Not Discussed (n=40)	Unclear (n=11)
Bias After Weighting				
Bias Reduction	85.6%	0.0%	77.5%	27.3%
No Bias Reduction	7.2%	66.7%	10.0%	0.0%
Not Discussed	4.5%	33.3%	10.0%	27.3%
Unclear	2.7%	0.0%	2.5%	45.5%

NOTE: 165 eligible studies were included in this analysis.

DISCUSSION

The present study provides the first systematic review of NRB studies involving federal surveys since the 2006 OMB *Standards and Guidelines for Statistical Surveys* were published. The review describes the general characteristics of NRB studies used in Federal surveys, the types of methods used to assess NRB, and the reported impact of mitigating post-survey adjustment strategies on final survey estimates.

The largest number of studies collected were conducted in the period immediately following the publication of the 2006 OMB *Standards and Guidelines for Statistical Surveys.*

The number of studies varied by federal agency with the largest proportion of studies coming from the NCES, BLS, and the NCHS.

Overall, about an equal number of household and establishment surveys were included in this review, and the various modes of data collection were represented – 61 paper surveys, 27 face-to-face surveys, 26 telephone surveys, and 45 surveys using multimode data collection strategies. The modes of data collection differed by survey type, with household surveys relying primarily on telephone, face-to-face, and multimode data collection approaches. On the other hand, the majority of establishment surveys used mail or school administered paper surveys only, or mail survey with telephone follow-up. Importantly, the type and mode of data collection appeared to affect the types of NRB methods employed to assess bias. For example, a study based on a household face-to-face survey with a paucity of frame information was apt to rely on examining variation in response by groups within the responding sample, whereas an establishment mail survey with considerable frame data is apt to assess survey estimates with reference to the frame information.

The process of gathering studies for this review highlights the marked difficulty of identifying NRB research involving federal surveys. As a result, we are limited in assessing the magnitude of the problem. Studies for this review were gathered through multiple avenues, including extensive literature searches and solicitations to federal statistical agencies and the greater statistical community. A large proportion of these studies were grey literature materials from government agencies, or the survey organizations collecting the data and were not widely published. But, despite our efforts, the collection is limited. Also, important, the reports included in this review lacked standardization and followed diverse reporting formats, making the extraction of data especially challenging. For example, the reporting of response rates varied

greatly between reports – nearly 20% of the studies included did not report a response rate, and most studies did not describe the specific standardized formulas used to calculate survey response rates.

Another challenging aspect of this review was ascertaining the potential for NRB in estimates prior to NR weight adjustments and the impact of those adjustments on reducing any bias in final survey estimates. There was a tendency in many studies for authors to declare that any bias discovered is not worthy of concern or attention. Furthermore, while the majority of studies reported a reduction in NRB on final estimates after weighting, this was simply asserted in most instances and not explicitly documented in the reports.

Since there is no current repository of all NRB studies for federal surveys, the observations made here are limited by the corpus of studies gathered for this systematic review. It is certainly possible that the collected sample may not represent the entire population of NRB studies conducted during this period. Without a centralized, common reporting approach across agencies, ensuring participation from researchers, a definitive analysis of studies is impossible to achieve.

POTENTIAL NEXT STEPS

The analysis presented here could be extended by redoubling efforts to gather studies from agencies whose work may be underrepresented. This may be an important, short term goal to improve upon the information already assembled. However, to better ensure that we can assess the nonresponse bias issue across federal surveys, a standardized approach to designing and reporting the findings of NRB studies and a centralized repository for them could be considered.

For example, agencies could be provided with a template to follow when conducting NRB studies and an online tool for submitting study details and findings.

Employing multiple NRB study methods, as recommended by Groves and Brick, could yield fuller pictures of the impact of nonresponse on potential bias in survey estimates. Investments in richer frame information for household surveys would facilitate more kinds of NRB analysis for this type of data collection. Conversely, establishment surveys could pursue methods beyond common analyses that focus on frame data. Recognizing that all NRB assessment methods have limitations, examining the problem from multiple angles seems to be a potentially fruitful approach. The FCSM could facilitate cross-agency conversations focused on how multiple NRB methods can be utilized in surveys with different data collection modes.

The movement within federal statistics to make use of alternative data sources does not lessen the importance of survey data. Surveys will continue to provide key benchmarks for estimates derived from administrative and unstructured data. Survey information increasingly will be combined with other data sources to construct desired estimates. The continued importance of survey data means that systematic monitoring of nonresponse bias in survey estimates is essential. This report offers information that may be useful in developing a centralized monitoring process.

SUPPLEMENTAL FILE I: Solicitations to AAPORnet and ASA listservs

AAPORnet:

In 2006, OMB issued guidelines concerning NRB analyses in federal surveys. The guidelines stipulated that plans for NRB analyses should be undertaken when surveys achieved a unit response rate below 80 percent or item nonresponse rates below 70 percent for items used in the report of survey findings.

A working group of the Federal Committee on Statistical Methodology is assembling a database of studies on NRB in federally sponsored surveys completed since the 2006 guidelines were issued. Our aim is to provide a picture of the extent and nature of NRB in federally sponsored surveys. We have conducted an extensive literature review thus far. We now appeal to members of the survey research community for help in identifying relevant studies. We will present a report on this project at the FCSM research and policy conference in March 2018.

We would greatly appreciate your help in identifying studies that should be included in the database. It is important for us to include conference papers and internal organization reports as well as published studies. We want to represent both studies conducted by federal agencies and ones conducted by other organizations for federally sponsored surveys.

Would you please send us links or references to NRB studies conducted since 2006 for federally sponsored surveys?

In order for responses to be useful, we need to receive them by **January 5, 2018**.

Please send links, references and any questions to this address: Kathryn.Piscopo@samhsa.hhs.gov

Thank you very much for your help.

Kind regards. Peter

ASA listserv:

In 2006, OMB issued guidelines concerning NRB analyses in federal surveys. The guidelines stipulated that plans for NRB analyses should be undertaken when surveys achieved a unit response rate below 80 percent or item nonresponse rates below 70 percent for items used in the report of survey findings.

A working group of the Federal Committee on Statistical Methodology is assembling a database of studies on NRB in federally sponsored surveys completed since the 2006 guidelines were issued. Our aim is to provide a picture of the extent and nature of NRB in federally sponsored surveys. We have conducted an extensive literature review thus far. We now appeal to members of the statistical community for help in identifying relevant studies. We will present a report on this project at the FCSM research and policy conference in March 2018.

We would greatly appreciate your help in identifying studies that should be included in the database. It is important for us to include conference papers and internal organization reports as well as published studies. We want to represent both studies conducted by federal agencies and ones conducted by other organizations for federally sponsored surveys. Studies should concern individual federally sponsored surveys conducted since 2006.

Would you please send us links or references to NRB analyses conducted since 2006 for federally sponsored surveys?

In order for responses to be useful, we need to receive them by January 5, 2018.

Please send links, references and any questions to this address: Kathryn.Piscopo@samhsa.hhs.gov

SUPPLEMENTAL TABLE I:

Table I. Number of non-response bias studies at each study phase

Study phase	Number of studies
After 2010 OMB solicitation	87*
After March 2017 literature search	23
After May 2017 literature search	59
After ASA/AAPOR solicitation	240*
Gathered post 2018 FCSM	16
Final included and coded studies	165
Final excluded studies	69

*Included some duplicate reports and non-pertinent documents.

SUPPLEMENTAL FILE II: NRB analysis codebook

NRB Study Identification Number: 001-999

NRB Study Title: Alphanumeric

NRB Study First Author: Alphanumeric

Year of Survey: 2006-17; Not provided = 99

Year of NRB Study: 2006-17

Agency Sponsor for Survey: 01-99

- ☐ Bureau of Economic Analysis = 01
- ☐ Bureau of Justice Statistics = 02
- ☐ Bureau of Labor Statistics = 03
- ☐ Bureau of Transportation Statistics = 04
- ☐ Economic Research Service = 05
- ☐ National Agricultural Statistics Service = 06
- ☐ National Center for Education Statistics = 07
- ☐ National Center for Health Statistics = 08
- ☐ National Center for Science and Engineering Statistics = 09
- ☐ Office of Research, Evaluation, and Statistics = 10
- ☐ Statistics of Income = 11
- ☐ US Census Bureau = 12
- ☐ US Energy Information Administration = 13
- ☐ Other = 98

Sponsor Agency Not Provided = 99

Response Rate: Single or Multiple (i.e. multi-year or single year multi-population response rates reported: 1 = Single; 2 = Multiple)

Unit Response Rate for Survey: 01-98; Not Provided = 99

(For multi-year surveys, code most recent RR; for single year, multi-population surveys – e.g. multi-state – code average or median RR if available; code 99 if not provided)

Survey Type:

11 = Household, General Population – e.g. Adults, 18-70 in United States

12 = Household, Special Population – e.g. VA hospital patients

18 = Household, Not Elsewhere Classified

21 = Establishment, non-farm, non-health care, non-school entities

22 = Establishment, Farm

23 = Establishment, School

24 = Establishment, Health Care Facility

28 = Establishment, Not Elsewhere Classified

Survey Mode

11 = Paper, Mail

12 = Paper, in person delivery (e.g. classroom administration)

21 = Telephone, RDD

22 = Telephone, List Sample

31 = Face-to-Face, e.g. CAPI

41 = Web, list sample

42 = Web, online panel

51 = Multimode survey (e.g. mail w phone follow up)

98 = Other

99 = Mode Not Provided

Target of NRB Analysis:

1 = Survey Estimates

2 = Sample Composition (Demographics);

3= Both Survey Estimates and Sample Composition

Reported NRB Prior to Weighting:

1 = Evidence of bias is reported;

5 = Evidence of no bias is reported;

7 = NRB prior to weighting is not discussed;

9 = Report is unclear.

Reported NRB After Weighting:

1 = Evidence of bias reduction in at least one variable is reported

5 = No evidence of bias reduction in at least one variable is reported

7 = NRB after weighting is not discussed;

9 = Report is unclear.

Method used in NRB Study:

11 = Study compared estimates from survey to estimates from other sources (benchmarking)

21 = Study compared estimates from survey to variables available on sampling frame

22 = Study compared estimates from survey to variables available on auxiliary data matched to sample

23 = Study compared estimates from survey to variables available from observations taken during data collection

24 = Study compared estimates from survey to variables available in seeded sample

31 = Study examined variation in response rates on subgroups

32 = Study examined variation within the respondent set, using information from prior wave data collection

33 = Study examined variation within the respondent set, following up on nonrespondents

34 = Study examined variation within the respondent set, through two phase (double) sampling of nonrespondents

35 = Study examined variation within the respondent set by analyzing estimates by level of effort 36 = Study examined variation within the respondent set by mounting randomized nonresponse experiments

41 = Study altered weighting adjustments, preparing estimates under different assumptions

42 = Study altered weighting adjustments, adjusting using models of characteristics
43 = Study altered weighting adjustments using models of response propensity

44 = Study altered weighting adjustments, adjusting using selection (Heckman) models

98 = Other method

99 = No mention of method

REFERENCES

Campbell, D. T. and D. W. Fiske (1959). "Convergent and discriminant validation by the multitrait-multimethod matrix." Psychological Bulletin **56**(2): 81-105.

Czajka, J. L. and A. Beyler (2016). Declining Response Rates in Federal Surveys: Trends and Implications (Background Paper), Mathematica Policy Research Reports, Mathematica Policy Research.

Federal Committee on Statistical Methodology. (2009). "How to Do Survey Nonresponse Bias Analyses in Household and Establishment Surveys.", from https://nces.ed.gov/FCSM/pdf/2009WorkshopProgram.pdf.

Groves, R. M. and J. M. Brick (2005). "Practical Tools for Nonresponse Bias Studies. Short Course Materials. Joint Program in Survey Methodology.".

Higgins, J. P. and S. Green (2008). "Cochrane handbook for systematic reviews of interventions."

Lin, I. F. and N. C. Schaeffer (1995). "Using Survey Participants to Estimate the Impact of Nonparticipation." The Public Opinion Quarterly **59**(2): 236-258.

Office of Management Budget. (2006). "Standards and Guidelines for Statistical Surveys." from https://georgewbush-whitehouse.archives.gov/omb/inforeg/statpolicy/standards_stat_surveys.pdf.

Seastrom, M. and National Center for Education Statistics (NCES) (2002). NCES Statistical Standards (NCES 2003–601). US Department of Education. Washington DC. **18**.

Teitler, J., N. Reichman and S. Sprachman (2003). "Costs and Benefits of Improving Response Rates for a Hard-to-Reach Population." Public Opinion Quarterly **67**.